7 November 2008 | $10

Science

Genetics of Behavior

AAAS

COVER

Honey bees (*Apis mellifera*), here shown on a honeycomb, form complex societies and interact with one another by means of stereotyped social behaviors. A special section beginning on page 891 explores what genetic approaches have taught us about behavior in bees and other species, including humans.

Image: Don Farrall, Getty Images

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

Genetics of Behavior

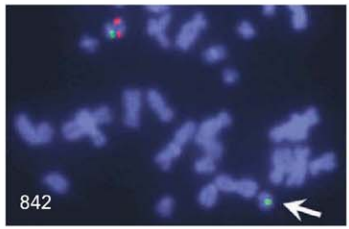
INTRODUCTION

NEWS

REVIEWS

PERSPECTIVES

>> Editorial p. 821, News story p. 842, Policy Forum p. 861, Reviews pp. 876 and 881; Science Podcast



NEWS OF THE WEEK

NEWS FOCUS

CONTENTS continued >>

SCIENCE EXPRESS

www.sciencexpress.org

MOLECULAR BIOLOGY

The *Air* Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin

T. Nagano et al.

Air, a large noncoding RNA, interacts with chromatin at a particular promoter, recruiting a histone methyltransferase to silence gene expression in an allele-specific manner.

10.1126/science.1163802

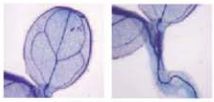
PLANETARY SCIENCE

Long-Lived Volcanism on the Lunar Farside Revealed by SELENE Terrain Camera

J. Haruyama et al.

Images of the Moon by the SELENE spacecraft and revised dates of lava flows by crater counts imply that episodic volcanism on the farside lasted to 2.5 billion years ago.

10.1126/science.1163382



MOLECULAR BIOLOGY

Photoexcited CRY2 Interacts with CIB1 to Regulate Transcription and Floral Initiation in *Arabidopsis*

H. Liu, X. Yu, K. Li, J. Klejnot, H. Yang, D. Lisiero, C. Lin

Blue light triggers the association of a photoreceptor, transcription factor, and DNA site, thus inducing expression of the gene *FT* (flowering time) and initiating flowering.

10.1126/science.1163927

LETTERS

TECHNICAL COMMENT ABSTRACTS

GENETICS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

ESSAY

REVIEWS

NEUROSCIENCE

GENETICS

BREVIA

BIOCHEMISTRY

RESEARCH ARTICLES

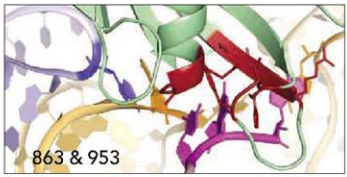
CELL BIOLOGY

REPORTS

APPLIED PHYSICS

CONTENTS continued >>

REPORTS *CONTINUED...*



AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) **is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005.** Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2008 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $970; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488322. Publications Mail Agreement Number 1069624. SCIENCE is printed on 30 percent post-consumer recycled paper. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 backissue prepaid (includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Printed on 30% post-consumer recycled paper.

CONTENTS continued >>

Not for sale.

SCIENCE**NOW**

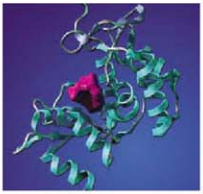
www.sciencenow.org

HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

River Dolphin Love Fetishes Not as Advertised
The boto has quite a reputation—but fortunately the body parts sold at Amazonian markets are fakes.

Diesel Fuel From a Tree Fungus?
Microbe's hydrocarbon stew could one day supplant fossil fuels.

Ancient Grave May Have Belonged to Shaman
Find hints at development of religion during crucial time in human history.



CD38, NAADP controller.

SCIENCE **SIGNALING**

www.sciencesignaling.org

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

REVIEW: NAADP—A Universal Ca^{2+} Trigger
A. H. Guse and H. C. Lee
NAADP elicits an initial release of calcium, which is subsequently amplified through the action of other calcium messengers.

ST NETWATCH: The Nobel Prize in Chemistry 2008
This year's award went to the scientists who discovered green fluorescent protein and developed it as an experimental tool; in Awards and Announcements.

ST NETWATCH: ASBMB Interactive Features
View Awards Lectures and interviews with prominent cell biologists from the American Society for Biochemistry and Molecular Biology; in Web Broadcasts.

ST NETWATCH: AMBMR Webcasts
View seminars from the annual meeting of the American Society for Bone and Mineral Research; in Web Broadcasts.

Lessons for postdocs.

SCIENCE **CAREERS**

www.sciencecareers.org/career_development

FREE CAREER RESOURCES FOR SCIENTISTS

Taken for Granted: Postdocs and Joe the Plumber
B. L. Benderly
What early-career scientists can learn from the working class.

In Person: The Trick to a Rewarding Career
M. Hermann
Throughout his career, Marc Hermann has always done exactly what he wanted.

U.K. Visa Changes Mean Closer Scrutiny for Non-European Students
A. Saini
New U.K. immigration policies impose more rigid requirements than in the past.

November 2008 Funding News
J. Fernández
Learn about the latest in research funding, scholarships, fellowships, and internships.

SCIENCE**PODCAST**

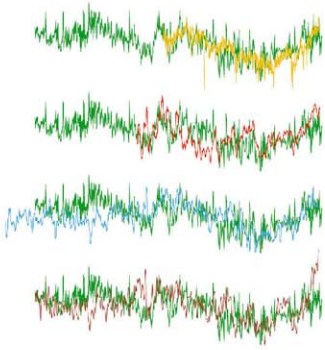
www.sciencemag.org/multimedia/podcast

FREE WEEKLY SHOW

Download the 7 November *Science* Podcast to hear about genes and social behavior, switching memories on and off, chromosome 17 and mental disorders, and more.

CREDITS: (SCIENCENOW) IZEN FARIAS/TOMAS HIRBEK; (SCIENCE CAREERS) ROBERT PODGORZELSKI/CREATIVE COMMONS; (SCIENCE SIGNALING) CHRIS BICKEL

Separate individual or institutional subscriptions to these products may be required for full-text access.



<< Tales of the Asian Monsoon

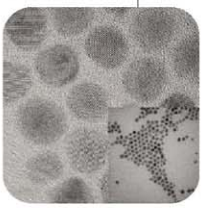
The Asian Monsoon is important for climate because it transports large amounts of heat and moisture from the ocean to the land. The monsoon is also important for human settlement because agriculture depends on monsoon rainwater. Using a record derived from a Chinese stalagmite, **Zhang *et al.*** (p. 940) present a detailed history of the Asian Monsoon over the past 1800 years that indicates connections between the monsoon, solar irradiation, Northern Hemispheric temperature, and glacial cycles in Europe. Shifts in the strength of the monsoon also correlate with the succession of Chinese dynasties, underscoring the importance that climate can have on human societies.

Stem Cells on Demand

Infection of adult mouse cells with viruses expressing genes of four transcription factors (Oct4, Sox2, c-myc, and Klf4) generates pluripotent stem cells (iPS) that resemble embryonic stem cells. Viruses commonly used for this procedure permanently alter the cells' genome and can cause tumors in animals, and thus these iPS cells cannot be used directly for cell therapy. **Stadtfeld *et al.*** (p. 945, published online 25 September) have produced mouse iPS cells by transiently exposing adult skin and liver cells to the four transcription factor genes using adenoviruses (that generally do not integrate into the genome). Thus, it is possible to make iPS cells without permanent genetic manipulation and it should be possible to make patient-specific cells not only to study disease but also for the future use of iPS cells in a clinical setting.

Sheltering Excitons in Quantum Dots

Quantum dots can exhibit long-lived fluorescence, but their excitonic states, which potentially are useful in photovoltaic and infrared detection applications, tend to decay very rapidly (in less than 1 picosecond). **Pandey and Guyot-Sionnest** (p. 929) report that the cooling of the two lowest energy excited states in cadmium-selenium quantum dots can be slowed by a thick coating of an electron insulator, in this case zinc-selenium. By using such insulation, the lifetimes of the excitonic states were extended to more than a nanosecond.



CREDITS (TOP TO BOTTOM): ZHANG *ET AL.*; PANDEY *ET AL.*

Going for the Burn

Fitness classes promote the idea that burning fat makes people healthier and perhaps live longer. **Wang *et al.*** (p. 957; see the Perspective by **Xie**) find that the *Caenorhabditis elegans* roundworm also adopts a fat-burning strategy to help them extend life. Up-regulation of a specific lipase, K04A8.5, decreases fat storage and increases life-span. The lipase level is low during adulthood but can be induced 10-fold when germline stem cells stop proliferating. In addition, the lipase contributes to longevity in worms by reducing insulin signaling. Thus, at least in *C. elegans*, fat metabolism and life-span control are directly linked.

Himalayan Graphite

Earth has an oxygen-rich atmosphere because, on a geological time scale, more organic material is created by photosynthesis than is respired back to carbon dioxide. Thus, knowing the particulars of how organic carbon is transformed by various geological processes, such as mountain-building, is essential for understanding the carbon and oxygen cycles. **Galy *et al.*** (p. 943) report that organic carbon is converted with surprising efficiency to graphite in the Himalayas, and subsequently buried in marine sediments after fluvial transport. During the erosion cycle up to half the carbon in the rocks was turned into graphite and sequestered in sediments, suggesting that the process could operate on a global scale to control carbon and oxygen cycles.

Moving Memories

The earliest phases of memory acquisition rely on the hippocampus, but growing evidence suggests that another area in the brain called the medial prefrontal cortex may take over in consolidated associative memories. **Takehara-Nishiuchi and McNaughton** (p. 960) found that following acquisition of an associative memory, neuronal activity in the rat medial prefrontal cortex became specific and necessary for the acquired memory. Selective activity patterns developed spontaneously during a consolidation period of about 6 weeks even without repetitive conditionings. Thus, a neural correlate of the memory gradually develops in the neocortex simultaneously with memory consolidation.

Mantle Flow

Seismic data provide an image of Earth's mantle today. Geologic data from mountain belts or sedimentary records in basins record the overall effects of mantle flow, but may not reveal the actual flow patterns. Starting with these observations, plus estimates of mantle properties, **Liu *et al.*** (p. 934; see the Perspective by **Steinberger**) have developed a model of the evolution of western North America during the past 100 million years. The model is consistent with flat subduction of the Farallon oceanic plate beneath the continent during much of this time, but shallow subduction extended over a larger area, which could explain a broad Cretaceous unconformity in sedimentary records.

Dry as a Bone Moon

Radar observations of the Moon from the Clementine spacecraft indicated water ice is present in the permanently shadowed Shackleton crater at the south pole. However, this finding has not been confirmed by Earth-based radar. Using a camera on board the SELENE (Kaguya) spacecraft, now orbiting the moon,

Continued on page 819

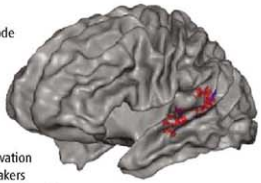
Continued from page 817

Haruyama *et al.* (p. 938) have been able to peer into this crater, and using the faint light reflected off the crater walls have measured the albedo and estimated surface temperatures across the interior. Although sufficiently cold (about 90 kelvin), the crater apparently lacks any large expanse of exposed ice. It is still possible that there are small amounts of ice beneath the surface, or mixed in the surface dust.

More Stem Cells on Demand

To rule out any risk of viral vectors integrating into the host genome and causing tumors, **Okita *et al.*** (p. 949, published online 9 October) used a plasmid transfection procedure to introduce transcription factor genes into mouse embryonic fibroblasts to make pluripotent cells. These cells show many features of embryonic stem cells, including the expression of pluripotency markers, as well as the capacity to develop teratomas and chimeras when transplanted into mice. Importantly, there was no evidence of plasmid integration and, although less efficient than other methods, this method looks like it will offer a safer way of inducing pluripotent stem cells.

Processing Speech and Voice



In everyday life, we automatically and effortlessly decode speech into language independently of who speaks. Similarly, we recognize a speaker's voice independently of what he says. **Formisano *et al.*** (p. 970) show that it is possible to decode the contents of speech and the identity of the speaker from measurements of the brain activity of a listener. They map and decode, trial-by-trial, the spatially distributed activation patterns evoked by listening to different vowels or speakers evoke in distinct patches of the listeners' auditory cortex. The pattern associated with a vowel does not change if the vowel is spoken by another speaker and the pattern associated with a speaker does not depend on what the person says.

CREDIT: FORMISANO ET AL.

Brain Repair

In mammals, a severed nerve in an arm or leg will eventually regrow and reestablish functional connections. A similar injury in the spinal cord or within the brain will not be repaired, resulting in permanent disability and paralysis. Poor regeneration in the central nervous system has been attributed to proteins embedded in brain myelin (the membranes that wrap each nerve axon), which interact with an inhibitory receptor on neurons called NgR. Two papers in this issue show that other inhibitory receptors recognize the myelin-embedded proteins (see the Perspective by **Kim and Snider**). **Atwal *et al.*** (p. 967) identified PirB, a mouse protein related to the immunoglobulins of the immune system, and if both PirB and NgR were blocked, regeneration resumed. **Park *et al.*** (p. 963) found that after injury to the optic nerve, the axons of the retinal ganglion cells in mice will regenerate if the growth-related signaling pathway mTOR is activated in these cells. When negative regulators of the mTOR pathway were deleted in the retinas of mice, within a few weeks, the axons of retinal ganglion cells would re-grow as far as the optic chiasm. Thus, to promote recovery from neural damage, a combination of therapeutic approaches is needed to remove inhibitory processes, as well as to stimulate the intrinsic growth pathways of the neurons.

Not Quite Sleep

Although thousands of people are made unresponsive under anesthesia, they are not always rendered unconscious, and stories of waking, eviscerated, on the operating table abound. **Alkire *et al.*** (p. 876) review what little we do know about the gap between behavioral unresponsiveness and oblivion. Although the relative role of the thalamus and cortical areas in switching consciousness on and off is not clear, despite their different mechanisms of action it does seem that most anesthetics hit a posterior corticothalamic complex centered around the inferior parietal lobe. As well as deactivating this region, anesthesia also causes functional disconnection between subregions of the complex. Understanding the effects of anesthesia could thus be a useful tool to understanding the neural correlates of consciousness.

Story Landis is director at the National Institute of Neurological Disorders and Stroke, National Institutes of Health.

Thomas R. Insel is director at the National Institute of Mental Health, National Institutes of Health.

The "Neuro" in Neurogenetics

THIS ISSUE OF *SCIENCE* FEATURES A SPECIAL SECTION (SEE PAGE 891) THAT FOCUSES ON AN emerging area of neurogenetics—the effort to link genomics and behavior. We are all intrigued by the notion that genomics may yield "simple" explanations for complex behaviors, including our own. The power of genomics has already revealed new insights into human disease and development. So what lessons has this new field taught us so far about behavior, and what can we look forward to in the coming decade?

One area of neurogenetics seeks the molecular basis for complex behaviors that range from mate choice in flies and social status in fish, to fidelity in voles and humans. Our intuition tells us that it should be easier to identify the mechanism underlying a simple reflex behavior (escape from threat) than a complex one (mate selection). But recent findings suggest that apparently simple genetic mechanisms may underlie some ostensibly complex behaviors. The field is just beginning to identify mechanisms for adaptive behaviors that are both parsimonious and profound. Although most research has investigated the genetics of behaviors in model organisms such as mice and flies, the diversity of the natural world is waiting to be mined. Behaviors that are unique to a species may be experiments of nature that can yield important insights into how genomic variation (inherited DNA sequence differences) relates to behavioral adaptation.

But perhaps most important in this burgeoning field is pursuit of the "neuro" in behavioral neurogenetics. Genes code for proteins, not for behaviors. By identifying how genomic variation modifies circuits of neurons, we will better understand both how and where behavior is instantiated. Most of the recently discovered variations are differences in the regulatory regions of genes that control gene expression. One important lesson from neurogenetics is that genomic variations in regulatory regions can account not only for how much of a protein is made, or when it is expressed, but exactly where in the brain a protein is expressed. Because brain function is specified by precise regional circuits, even small differences in the location of the brain cells that produce a particular receptor or an enzyme can result in large differences in function. Importantly, the link between genomic sequence and behavior is the brain: We cannot hope to understand how genomic variation influences behavior without understanding how genomic variation influences neural circuitry.

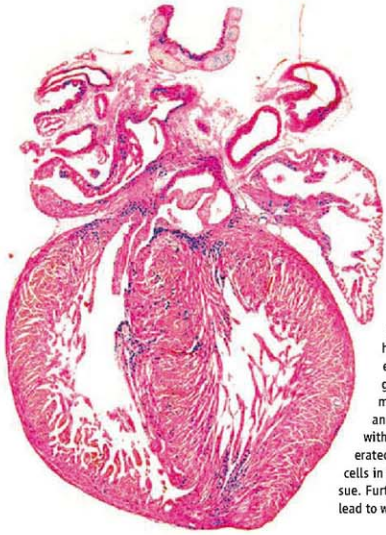
An early mainstay of neurogenetics was the identification of mutations in mice with abnormal behaviors. Here, discrete brain changes could be tied to regional molecular mechanisms once the genes were cloned. For human neurological disorders like Huntington's and Parkinson's disease, such obvious structural and functional changes can indeed explain how gene defects give rise to certain behaviors. For psychiatric diseases, where the neurobiological lesions are not known, the challenge will be greater. But genomics now promises to be the key to unlock the neurobiology of these complex disorders: There is real hope that genomic variation will lead us to neural mechanisms that can begin to explain such complex syndromes as schizophrenia or autism.

Behavioral phenotypes are the result of a complex interaction between nature (DNA) and nurture (experience). The developing brain is the stage for this drama, but we still know little about the details of how genes and experience interact within the developing brain to create something as complex as the phenotype we call human nature. In the coming decades, epigenomics—changes in gene expression due to alterations in protein-DNA interactions rather than DNA sequence—will be critical for understanding how experience alters the genome, complementing the current focus on how genomic variation affects behaviors.

Already it is clear that, for the study of behavior, genomics is not destiny. Indeed, if genomic sequence "determines" anything behaviorally, it determines diversity. It is important that we be wary about extrapolating from model organisms to humans. We must also avoid using small statistical associations to make grand claims about human nature. Obviously, we have much to discover before understanding how genes influence behavior—a discovery process that will closely involve the brain.

–Story Landis and Thomas R. Insel

10.1126/science.1167707



DEVELOPMENT

Healing a Broken Heart

The ability to regenerate damaged tissues and organs varies widely across animals. While mammals are able to repair ruptured muscles and to regrow fingertips, amphibia and fish have the more resilient tissues, being able to regenerate tails, fins, and even hearts. Although heart regeneration was thought to be restricted to a few species of amphibia, it is of particular interest to humans, because coronary heart disease remains a leading cause of death. Drenckhahn *et al.* have found that the fetal mouse heart is able to replace damaged tissue. The enzyme holocytochrome c synthase (Hccs) is involved in mitochondrial energy generation, and the authors inactivated the X-linked *Hccs* gene in female mice. At mid-gestation, heterozygous female hearts contained equal numbers of healthy and damaged cells; by the time of birth, these mice had fully functioning hearts, with less than 10% damaged cells. Thus, the mouse fetal heart appears to be regenerated predominantly from differentiated cardiac cells, suggesting that differentiated cells in the adult might retain an intrinsic capacity to expand and replace damaged tissue. Further studies aimed at understanding the molecular mechanisms involved could lead to ways of stimulating the regeneration of adult diseased hearts. — HP*

Dev. Cell **15**, 521 (2008).

IMMUNOLOGY

Lymphocyte Identity Cards

The determination of lineage, whether in genealogy, paleontology, or cell biology, can be very difficult. Schepers *et al.* have developed a retroviral tagging procedure by introducing a "bar code" into individual cells that persists in all of their progeny. The authors used a library of around 5000 tags, which can be identified by PCR amplification and microarray analysis, to monitor the life histories of T cells during the course of an infection.

A T cell population, specific for the antigen OVA, was transformed with the bar-code library and introduced into mice, which were subsequently injected with tumor cells and infected with influenza virus, both bearing the OVA antigen. At first, T cells in lymph nodes draining the two sites of invasion formed genetically distinct populations, distinguishable by their bar codes; however, the T cell populations in lung and tumor tissues had similar bar-code distributions, showing that they had originated from several lymph nodes. Over time, both lymph node T cell populations became similar as the infections stimulated the migration of T cells throughout the mouse. This technology has the potential to unravel lineage relationships in a wide range of cells, and the authors have already created a lentivirus library for use with quiescent cell types resistant to retroviral infection. — CS*

J. Exp. Med. **205,** 2309 (2008).

CREDITS (TOP TO BOTTOM): DRENCKHAHN *ET AL.*, *DEV. CELL* 15, 521 (2008); DEMETRIOU *ET AL.*, *PHYS. REV. LETT.* 101, 145702 (2008)

*Helen Pickersgill and Chris Surridge are locum editors in *Science*'s editorial department.

MATERIALS SCIENCE

Sizing Up the Foam

Bulk metallic glasses have high plastic yield strengths, and thus have the potential for making ultrastrong foams. However, the foam will only inherit the strength of the parent glassy material if it fails by plastic yielding, rather than by brittle fracture (which is associated with the solid fracture stress) or by elastic buckling (associated with the solid modulus). Demetriou *et al.* look at a number of critical structural scales that influence the failure mode and find that they can make ultrastrong glassy foams from a $Pd_{43}Ni_{10}Cu_{27}P_{20}$ alloy with up to 92% porosity. The foams were engineered against buckling and fracture though a process that limited membrane thickness and promoted cellular periodicity. Evaluation of compressed, collapsed specimens showed both crushed cells and shear banding, indicating that although the failure was due to fracture, the initial response of the foam involved plastic deformation. Thus, the foams inherited the best properties of the parent glassy material. The compressive strength of the glassy foams rivaled those obtained for highly engineered Ti-6Al-4V or ferrous metal foams. — MSL

Metallic-glass foam.

Phys. Rev. Lett. **101,** 145702 (2008).

SYSTEMS BIOLOGY

Network Failure

Models of metabolic and signalling networks have been characterized, perhaps unfairly, as reannotations of previously discovered interactions. To counter this concern (and the statistical issue of sorting through hundreds of correlations), Janes *et al.* describe an approach called "model breakpoint analysis" that stresses the network by using non-physiological inputs in a manner similar to that of

Continued on page 825

Continued from page 823

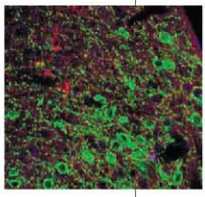
engineers performing failure analysis of bridges or cars. They began with their model of cytokine-induced apoptosis and proceeded to introduce implausible data that stretched the dynamic range of the cell (defined as the responsiveness of cell outcomes to incremental changes in cell activation). Surprisingly, network function did not degrade in parallel, but worked perfectly well until a threshold (or breakpoint) was reached, at which point the predictions were no longer useful. Pinpointing the signals and stimuli that were responsible for the system failure enabled them to distinguish epiphenomena from causal factors and to make predictions about the dynamic roles of three kinases (Akt, ERK, and Mk2) in cytokine-induced apoptosis. These predictions were then confirmed in inhibitor- and mutant-based experiments, suggesting that differences in dynamic range can be more important to cellular function than the strength of a particular signal. — BJ

Cell **135**, 343 (2008).

NEUROSCIENCE

Depotentiating via Dopamine

The capacity to associate events, in a neuronal context, is thought to rely on long-term potentiation (LTP), a mechanism that strengthens glutamatergic (excitatory) synaptic connections. Strong novel stimuli can selectively reverse or overwrite LTP by a mechanism known as depotentiation, which is thought to keep synapses from becoming saturated and thereby to maintain them in a dynamically responsive range. Neuregulin-1 is a factor expressed in brain and can effectively depotentiate LTP in the hippocampus. Kwon *et al.* found that neuregulin depotentiates LTP by recruiting a dopaminergic signaling pathway involving the dopamine D4 receptor (D4R), which is a target of the antipsychotic clozapine. Neuregulin acutely triggers dopamine release in the hippocampus, which in turn depotentiates LTP by activating D4Rs. The direct activation of D4Rs by selective agonists mimics the action of neuregulin in removing AMPA-type glutamate receptors from synapses. Mutant mice lacking D4Rs fail to depotentiate LTP in response either to neuregulin or to electrical stimuli. These observations thus functionally associate three signaling pathways (dopamine, glutamate, and neuregulin) in the regulation of synaptic plasticity. — PRS



Dopaminergic (green) neurons in the ventral tegmental area.

CREDIT: KWON *ET AL.*, *PROC. NATL. ACAD. SCI. U.S.A.* 105, 15587 (2008)

Proc. Natl. Acad. Sci. U.S.A. **105**, 15587 (2008).

CLIMATE SCIENCE

Warming Vapors

Water vapor is the atmospheric gas that collectively has the greatest greenhouse effect on climate, although it does not directly instigate warming or cooling trends, because the amount of water vapor in the atmosphere varies only in response to temperature change. Instead, water vapor only amplifies temperature trends being caused by other factors such as atmospheric CO_2 concentration or Earth's albedo. The extent to which humidity changes in response to temperature variation is therefore a key parameter in global climate models, because that quantity determines the strength of the associated warming or cooling. Dessler *et al.* present satellite data from 2003 to 2008 which show that models have gotten that relationship correct, and that relative humidity is effectively constant at any given temperature. Thus, the temperature increases predicted by global models are virtually guaranteed to be several degrees Celsius by the year 2100. Knowing the water vapor content of a warmer atmosphere is also important for predicting rainfall and storminess. — HJS

Geophys. Res. Lett. **35**, L20704 (2008).

CELL BIOLOGY

Curing Disease in Yeast

Batten disease is a neurodegenerative disorder linked to the pathological accumulation of material in lysosomes. In yeast the [URE3] phenotype is a prion (infectious protein) generated by the self-propagating amyloid form of the Ure2 protein, which regulates nitrogen catabolism. Yeast prions can arise and disappear spontaneously within populations, reflecting in part changes in the protein folding milieu. Kryndushkin *et al.* show that increased production of Btn2 protein or its homolog Cur1 can cure [URE3]. Conversely, deletion of *BTN2* and *CUR1* genes stabilizes the [URE3] phenotype. In cells expressing a fluorescently tagged version of Btn2p, fluorescence accumulated at a single point close to the nucleus and vacuole, where aggregates of Ure2p also accumulated. This accumulation of protein aggregates reduced the ability of the Ure2p amyloid seeds to enter budding daughter cells, explaining the cure of daughter cells. This accumulation of protein aggregates mirrors aggresome formation observed in mammalian cells, which may also function to remove potentially harmful protein aggregates. — SMH

EMBO J. **27**, 2725 (2008).

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES please call 202-326-6755 for any questions or information

REPRINTS: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

PERMISSIONS 202-326-7074, FAX 202-682-0816

MEMBER BENEFITS AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store www.aplsource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/epptoaa/aaac; Bank of America MasterCard 1-800-833-4262 (priority code FAA3YU); Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Soabery & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

INFORMATION FOR AUTHORS
See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Bruce Alberts**
EXECUTIVE EDITOR **Monica M. Bradford**
DEPUTY EDITORS **R. Brooks Hanson, Barbara R. Jasny, Katrina L. Kelner**
NEWS EDITOR **Colin Norman**

EDITORIAL SUPERVISORY SENIOR EDITOR Phillip D. Szuromi; **SENIOR EDITOR/PERSPECTIVES** Lisa D. Chong; **SENIOR EDITORS** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Valda Vinson; **ASSOCIATE EDITORS** Kristen L. Mueller, Jake S. Yeston, Laura M. Zahn; **ONLINE EDITOR** Stewart Wills; **ASSOCIATE ONLINE EDITORS** Robert Frederick, Tara S. Marathe; **WEB CONTENT DEVELOPER** Martyn Green; **BOOK REVIEW EDITOR** Sherman J. Suter; **ASSOCIATE LETTERS EDITOR** Jennifer Sills; **EDITORIAL MANAGER** Cara Tate; **SENIOR COPY EDITORS** Jeffrey E. Cook, Cynthia Howe, Harry Jach, Barbara P. Ordway, Trista Wagoner; **COPY EDITORS** Chris Filstrup, Lauren Kmec; **EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields; **PUBLICATIONS ASSISTANTS** Ramatoulaye Diop, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Scott Miller, Jerry Richardson, Jennifer A. Seibert, Brian White, Anita Wynn; **EDITORIAL ASSISTANTS** Carlos L. Durham, Emily Guise, Patricia M. Moore; **EXECUTIVE ASSISTANT** Sylvia S. Kihara; **ADMINISTRATIVE SUPPORT** Maryrose Madrid

NEWS DEPUTY NEWS EDITORS Robert Coontz, Eliot Marshall, Jeffrey Mervis, Leslie Roberts; **CONTRIBUTING EDITORS** Elizabeth Culotta, Polly Shulman; **NEWS WRITERS** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin, David Grimm, Constance Holden, Jocelyn Kaiser, Richard A. Kerr, Eli Kintisch, Andrew Lawler (New England), Greg Miller, Elizabeth Pennisi, Robert F. Service (Pacific NW), Erik Stokstad, **NEWS** Rachel Ze!kowitz; **CONTRIBUTING CORRESPONDENTS** Jon Cohen (San Diego, CA), Daniel Ferber, Ann Gibbons, Robert Koenig, Mitch Leslie, Charles C. Mann, Virginia Morell, Evelyn Strauss, Gary Taubes; **COPY EDITORS** Linda B. Felaco, Melvin Gaither, Melissa Rahmondi; **ADMINISTRATIVE SUPPORT** Sherrie Sherraine Mack, Fannie Groom; **AMERICAS** New England: 207-549-7755, San Diego, CA: 760-942-3252, FAX 760-942-4979, Pacific Northwest: 503-963-1940

PRODUCTION DIRECTOR James Landry; **SENIOR MANAGER** Wendy K. Shank; **ASSISTANT MANAGER** Rebecca Doshi; **SENIOR SPECIALISTS** Steve Forrester, Chris Redwood; **SPECIALIST** Anthony Rosen; **PREFLIGHT DIRECTOR** David M. Tompkins; **MANAGER** Marcus Spiegler

ART DIRECTOR Yael Kats; **ASSOCIATE ART DIRECTOR** Laura Creveling; **ILLUSTRATORS** Chris Bickel, Katharine Sutliff; **SENIOR ART ASSOCIATES** Holly Bishop, Preston Huey, Nayomi Kevithyagala; **ART ASSOCIATE** Jessica Newfield; **PHOTO EDITOR** Leslie Ritzaert

SCIENCE INTERNATIONAL

EUROPE (science@science-int.co.uk) **EDITORIAL: INTERNATIONAL MANAGING EDITOR** Andrew M. Sugden; **SENIOR EDITOR/PERSPECTIVES** Julia Fahrenkamp-Uppenbrink; **NEWS EDITORS** Caroline Ash, Stella Hurtley, Ian S. Osborne, Peter Stern; **ASSOCIATE EDITOR** Marla Cruz; **EDITORIAL SUPPORT** Deborah Dennison, Rachel Roberts, Alice Whaley; **ADMINISTRATIVE SUPPORT** John Cannell, Janet Clements; **NEWS: EUROPE NEWS EDITOR** John Travis; **DEPUTY NEWS EDITOR** Daniel Clery; **CONTRIBUTING CORRESPONDENTS** Michael Balter (Paris), John Bohannon (Vienna), Martin Enserink (Amsterdam and Paris), Gretchen Vogel (Berlin); **INTERN** Sara Coelho

ASIA Japan Office: Asca Corporation, Eiko Ishioka, Fusako Tamura, 1-8-13, Hirano-cho, Chuo-ku, Osaka-shi, Osaka, 541-0046 Japan; +81 (0) 6 6202 6272, FAX +81 (0) 6 6202 6271; asca@os.gulf.or.jp; **ASIA NEWS EDITOR** Richard Stone (Beijing; rstone@aaas.org); **CONTRIBUTING CORRESPONDENTS** Dennis Normile (Japan: +81 (0) 3 3391 0630, FAX +81 (0) 3 5936 3531; dnormile@gol.com); Hao Xin (China: +86 (0) 10 6307 4439 or 6307 3676, FAX +86 (0) 10 6307 4358; cindyhao@gmail.com); Pallava Bagla (South Asia: +91 (0) 11 2271 2896; pbagla@vsnl.com)

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner**

FULFILLMENT SYSTEMS AND OPERATIONS (membership@aaas.org): **DIRECTOR** Waylon Butler; **SENIOR SYSTEMS ANALYST** Jonny Blaker; **CUSTOMER SERVICE SUPERVISOR** Pat Butler; **SPECIALISTS** Latoya Casteel, LaWonda Crawford, Vicki Linton, April Marshall; **DATA ENTRY SUPERVISOR** Cynthia Johnson; **SPECIALISTS** Elntou Bowden, Tamika Hill, William Jones

BUSINESS OPERATIONS AND ADMINISTRATION DIRECTOR Deborah Rivera-Wienhold; **ASSISTANT DIRECTOR, BUSINESS OPERATIONS** Randy Yi; **MANAGER, BUSINESS ANALYSIS** Michael LoBue; **MANAGER, BUSINESS OPERATIONS** Jessica Tierney; **FINANCIAL ANALYSTS** Priti Parmar, Celeste Trotter; **RIGHTS AND PERMISSIONS ADMINISTRATOR** Emilie Daniel; **ASSOCIATE** Elizabeth Sandler; **MARKETING DIRECTOR** John Meyers; **MARKETING MANAGER** Allison Pritchard; **MARKETING ASSOCIATES** Aimee Aponte, Allison Chandler, Mary Ellen Crowley, Marcia Leach, JoRanne Wielga, Wendy Wise; **INTERNATIONAL MARKETING MANAGER** Wendy Sturley; **MARKETING EXECUTIVE** Jennifer Reeves; **MARKETING/NUMBER SERVICES EXECUTIVE** Tosha Rich; **DIRECTOR, SITE LICENSING** Tom Ryan; **DIRECTOR, CORPORATE RELATIONS** Ellen Bernadette Moran; **PUBLISHER RELATIONS, eRESOURCES SPECIALIST** Kiki Forsythe; **SENIOR PUBLISHER RELATIONS SPECIALIST** Catherine Holland; **PUBLISHER RELATIONS, EAST COAST** Phillip Smith; **PUBLISHER RELATIONS, WEST COAST** Phillip Tsolakidis; **FULFILLMENT SUPERVISOR** Iqbo Edim; **ELECTRONIC MEDIA MANAGER** Elizabeth Harman; **PROJECT MANAGER** Trista Snyder; **ASSISTANT MANAGER** Lisa Stanford; **SENIOR PRODUCTION SPECIALISTS** Christopher Coleman, Walter Jones; **PRODUCTION SPECIALISTS** Michele Johnston, Kimberly Oster

ADVERTISING DIRECTOR, WORLDWIDE Ad Sales Bill Moran

PRODUCT (science_advertising@aaas.org): **MIDWEST/WEST COASTW. CANADA** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189; **UK/EUROPE/ASIA** Roger Goodwin: TEL/FAX +41 43 234 1358; **JAPAN** Mashy Yoshikawa: +81 (0) 3 3235 5961, FAX +81 (0) 3 3235 5852; **SENIOR TRAFFIC ASSOCIATE** Delandra Simms

COMMERCIAL EDITOR Sean Sanders: 202-326-6430

PROJECT DIRECTOR, OUTREACH Brianna Blaser

CLASSIFIED (advertise@sciencecareers.org): **US: RECRUITMENT SALES MANAGER** Ian King: 202-326-6528, FAX 202-289-6742; **INSIDE SALES MANAGER: MIDWEST/CANADA** Daryl Anderson: 202-326-6543; **INSIDE SALES REPRESENTATIVE** Karen Foote: 202-326-6740; **KEY ACCOUNT MANAGER** Jorbah Able; **NORTHEAST** Alexis Fleming: 202-326-6578; **SOUTHEAST** Tina Burks: 202-326-6577; **WEST** Nicholas Hinthlidze: 202-326-6533; **SALES COORDINATORS** Erika Foard, Rohan Edmonson, Shirley Young; **INTERNATIONAL: SALES MANAGER** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **SALES** Susanne Kharraz, Dan Pennington, Alex Palmer; **SALES ASSISTANT** Louise Moore; **JAPAN** Mashy Yoshikawa: +81 (0) 3 3235 5961, FAX +81 (0) 3 3235 5852; **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins; **SENIOR PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle; **SENIOR PRODUCTION SPECIALIST** Robert Buck; **SENIOR TRAFFICASSOCIATE** Christine Hall; **PUBLICATIONS ASSISTANT** Mary Lagnaoui

AAAS BOARD OF DIRECTORS RETIRING PRESIDENT, CHAIR David Baltimore; **PRESIDENT** James J. McCarthy; **PRESIDENT-ELECT** Peter C. Agre; **TREASURER** David E. Shaw; **CHIEF EXECUTIVE OFFICER** Alan I. Leshner; **BOARD** Lynn W. Enquist, Susan M. Fitzpatrick, Alice Gast, Linda P. B. Katehi, Nancy Knowlton, Cherry A. Murray, Thomas D. Pollard, Thomas A. Woolsey

Eyeing Balls

Useful news for tennis pros: Umpires are much more likely to make mistakes when calling balls out rather than in.

Scientists at the University of California, Davis, analyzed 4457 points from tennis matches played during the 2007 Wimbledon tournament in the U.K., including all challenges submitted by players. Of 83 recorded blunders, 70 were wrongly called out and only 13 wrongly called in. The skew is due to a perceptual bias toward the direction of movement of a bouncing ball, the authors reported in the 28 October issue of *Current Biology*. On top of that, says lead author David Whitney, umpires are more likely to "mislocalize" balls that are traveling toward them.

Tennis players can take advantage of this bias, he says, by concentrating their challenges on rulings that their own balls are "out"—rather than on rulings of "in" for an opponent.

Psychologist Alan Johnston of University College London says this perception bias "might have implications for other sports, such as soccer or rugby," for which player positions are at issue.

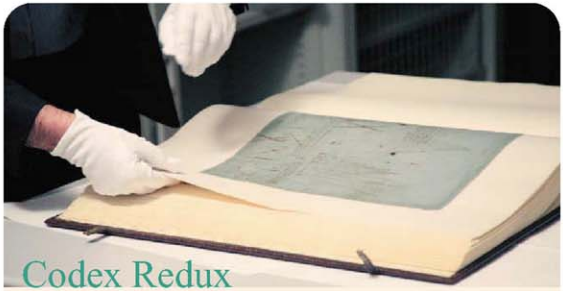
Gender and the Brain

The largest ever genetic study of male-to-female transsexuals has provided a hint—albeit a faint one—as to how gender is embedded in the brain. A team led by molecular geneticist Vincent Harley of the Monash Medical Centre in Melbourne, Australia, analyzed versions of

CREDITS (TOP TO BOTTOM): BIBLIOTECA AMBROSIANA, MILAN; DAVID PICKOFF/AP; EPFL

three hormone-related genes in 112 white male-to-female transsexuals recruited in Melbourne and Los Angeles, California. The findings were compared with DNA samples from 258 nontranssexual males. Categorizing the alleles as either "short" or "long," they found that the transsexuals had more long alleles for the androgen receptor gene, they reported online last week in *Biological Psychiatry*. Longer alleles, they explain, inhibit receptor activity, leading to less effective prenatal testosterone signaling. Although the effect is weak—55% of the transsexuals had the long allele, compared with 47% of the controls—the researchers suggest it could play a role in incomplete masculinization of the brain during early development.

Psychologist Kenneth Zucker, head of the Gender Identity Service at the Centre for Addiction and Mental Health in Toronto, Canada, says it's hard to see how prenatal hormones could affect brain development in that way without altering the sex organs as well. That's why "everybody is looking for some [other] type of marker," he says. Nonetheless, behavioral neuroendocrinologist Marc Breedlove of Michigan State University in East Lansing says "it will be exciting" if the finding is replicated.



Codex Redux

The 500-year-old Codex Atlanticus, a compilation of notes and drawings by Leonardo da Vinci, may be disassembled for better conservation, officials at Ambrosiana library in Milan, Italy, announced last week.

The Codex dates from the late 16th century, when an Italian sculptor gathered some 1120 pages of notes and drawings into a 402-page volume. In the early 1970s, restorers glued the notes onto blank sheets and split them into 12 books—a move that experts now say weakened the paper, altered edges, and made the pages awkward to display.

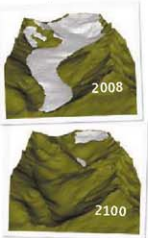
The Codex drew renewed attention last year when scholars noticed black stains on the support pages that they feared were from mold. After a year of research conducted at the Istituto Centrale di Patologia del Libro (ICPL), epidemiologist Gianfranco Tarsitani of the University of Rome "La Sapienza" announced last month that the stains were due not to mold but to mercury salts added as a preservative.

Leading da Vinci expert Carlo Pedretti, a professor emeritus at the University of California, Los Angeles, has given his blessing to the plan, which he says will help tailor conservation to the needs of individual pages. "Each time we have to study or display one drawing, we have to manipulate the whole volume," notes ICPL conservator Armida Batori.

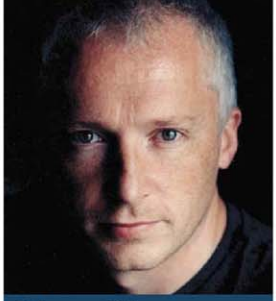
Running Out of Glacier Time

An iconic feature of the Swiss landscape may vanish within the century. Swiss researchers have developed a model simulating the retreat of the Rhône Glacier in southern Switzerland since 1874 that predicts the glacier's possible disappearance by 2100.

Mathematician Guillaume Jouvet of the Federal Polytechnic Institute in Lausanne, Switzerland, and colleagues tweaked a classic fluid-dynamics model to account for the viscosity of ice and accumulation of snow on the glacier, they report in an upcoming issue of the *Journal of Glaciology*. They also fed in more than a century of detailed temperature and precipitation data obtained from the Swiss Federal Institute of Technology Zurich (ETH).



If average global temperature goes up by 1°C by 2100, the glacier will lose about 35% of its volume, says ETH glaciologist Matthias Huss. Under the worst-case scenario, a nearly 4°C increase, the glacier will disappear completely by then. "In all possibilities, we have retreat," Huss says, with enormous consequences for water supplies and power generation as well as for ecology.



Three Q's >>

Acclaimed mathematician, best-selling author, newspaper columnist, and host of the current BBC television series *The Story of Maths*, **Marcus du Sautoy** has been appointed to the Simonyi Professorship Chair for the Public Understanding of Science at the University of Oxford in the U.K. He will succeed evolutionary biologist Richard Dawkins, who retired from the chair last month.

Q: Dawkins was a very high-profile holder of this professorship. What might you do differently?

I want to steer the position back to science rather than talking about religion, which is what I think it's been slightly concentrated on the past few years.

Q: Is communicating mathematics to the public more challenging than communicating evolution?

It is, partly because many of the things we study just exist in the mind and don't have a physical reality. But mathematics is fundamental to all the sciences. It is the language of nature.

Q: "Public understanding" is a difficult thing to quantify. How will you measure the success of your efforts?

Some indications [might be that] on [BBC] Radio 1, they're now quite happy to mention what a prime number is without batting an eye. That would not have been true 15 years ago. Once you see mathematics getting embedded into the public psyche and popular culture, that will be an indication that we're getting the message through. If we see fewer people saying "I hate maths" and actually choosing maths as a university subject, that will also be considered a success.

CREDITS (TOP TO BOTTOM): NIALL MCDIARMID; ELINCOLN ELSE, UNIVERSITY OF LETHBRIDGE

Got a tip for this page? E-mail people@aaas.org

MOVERS

NEW DIRECTION. DESY, Germany's particle physics lab near Hamburg, this week tapped a prominent solid-state physicist, Helmut Dosch of the Max Planck Institute for Metals Research in Stuttgart, to be its new director-general beginning in March 2009. This mismatch of disciplines reflects a shift in the lab itself. DESY's main accelerator, HERA, shut down in 2007, and the lab's main focus is now XFEL, an x-ray–free electron laser for studying the structure of matter that will be completed by 2013. "DESY will shed light on so-far-unexplored dimensions in nanospace," says Dosch.

WARM CASH. Lured by $15.7 million from the Alberta, Canada, government and the University of Lethbridge (UL), neuroscientist Bruce McNaughton is reestablishing his Canadian roots. Last month, the 60-year-old researcher at the University of Arizona, Tucson, joined UL's Canadian Centre for Behavioural Neuroscience and brought with him a talented former postdoc, David Euston, also from UA. More will follow, says McNaughton, who expects to provide a "kind of theoretical and computational perspective" for the center while steering younger scientists along promising avenues of research.

McNaughton, who left Canada in 1982 to do a postdoc in Norway, says he was attracted by the lack of strings attached to the prize, the first of three Polaris Investigator Awards that the province is offering to top-flight scientists from around the world (*Science*, 6 April 2007, p. 29).

McNaughton's only lament is the climate. "Ideally, one would live in Arizona at this time of year and come up here in the summertime," he says.

BACK TO THE LAB. J. Michael Bishop will step down in June 2009 as chancellor of the University of California, San Francisco (UCSF). The 72-year-old Nobelist will remain on the faculty as a professor of microbiology and immunology.

In his decade-long tenure as chancellor, Bishop oversaw both the construction of a second campus that will become one of the country's largest biomedical research centers and the establishment of an institute for research on stem cells and regenerative medicine. There were also some rough patches, among them lingering financial fallout from a failed hospital merger with Stanford University in Palo Alto in 1997 and the dismissal last year of medical school dean David Kessler over a dispute about the university's finances (*Science*, 21 December 2007, p. 1855).

Harold Varmus, who shared the 1989 Nobel Prize in physiology or medicine with Bishop, says Bishop has had a successful run as chancellor. "He enjoyed raising money and was enormously good at it at a time when UCSF was growing at a dramatic pace," says Varmus, adding that the university's reputation "just continues to improve."

PLUCKED. How often does a geologist have to be whisked off by helicopter in the middle of fieldwork? It happened to Greg Stock last month.

The staff geologist at Yosemite National Park in northern California was halfway into a 6-day climb to map the rock types on the face of the El Capitan mountain when he heard a loud rumble over his radio. A few kilometers away, a rockfall had sent nearly 6000 m^3 of stone tumbling down to Curry Village, a collection of rustic cabins and tent sites.

Slice of Life

The rockfall had slightly injured three people, and park managers wanted Stock to come over immediately to assess the chance of another incident. But he was several hundred meters up in the air, clinging to the world's most famous stone wall. So a rescue team landed a helicopter on top of the cliff, hauled Stock up with a long rope, and flew him to the accident site.

Ironically, Stock was climbing to learn more about why rockfalls happen. The previous night, he had heard of a much smaller rockfall at the accident site that "unsettled" him. Had he been able to check it out, it might have provided clues that the larger one was coming. "I wish that I wasn't up [on El Capitan] at the time," he says. The mapping effort, which Stock's guides completed, will help him locate the source of a mysterious 2.7-million-m^3 rockfall that came off El Cap 3600 years ago.

Climate and society in a Chinese cave
837

Science and the meltdown
841

NATIONAL INSTITUTES OF HEALTH

Zerhouni's Parting Message: Make Room for Young Scientists

An intractable problem faced Elias Zerhouni when he became director of the National Institutes of Health (NIH) 6 years ago: The agency's corps of more than 20,000 independent investigators was getting old. The average age at which researchers receive their first NIH research grant had been creeping up for decades. (It is now 42.) Zerhouni saw this as a crisis and tackled it head on. After probing the data, he launched an experiment. Instead of relying solely on peer review to apportion grants, he set a floor—a numerical quota—for the number of awards made to new investigators in 2007 and 2008.

Last week on his final day as director, Zerhouni made this a formal NIH policy. He hopes his successors will keep it: "I think anybody who thinks this is not the number-one issue in American science probably doesn't understand the long-term issues," he says. The notice states that NIH "intends to support new investigators at success rates comparable to those for established investigators submitting new applications." In 2009, that will mean at least 1650 awards to new investigators for R01s, NIH's most common research grant.

The quotas have meant pain for some institutes in a time when NIH's budget isn't growing. Many are trying to steer money to new grantees by setting funding cutoff points in peer-review scores at more generous levels for new investigators than for established ones. Although some scientists may see this as a kind of affirmative action, Zerhouni says it is not. To him, it is simply "leveling the playing field" by correcting peer reviewers' bias against the young.

In 1980, the average age of a first-time NIH grant recipient was 37. The 5-year rise in average age since then, observers say, can be blamed on longer time spent in training, including in postdocs, and the older age at which faculty are first hired at medical schools, where they begin independent careers. In 2003, when NIH's budget stopped growing, the situation "collapsed," Zerhouni says: The number of R01-like research grants (known as R01 equivalents) going to first-time investigators slipped to 1354 in 2006, the lowest level in 9 years.

Transfusion. Departing NIH Director Elias Zerhouni says it is urgent to bring new blood into U.S. biomedical research.

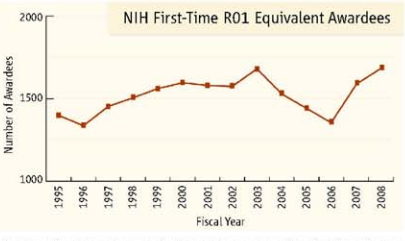
This is "detrimental for all sorts of reasons," says Jeremy Berg, director of the National Institute of General Medical Sciences. One concern is that scientists are not getting enough support when they're young, during their most creative years. Another is that the well may run dry. When Zerhouni asked his staff to model the age distribution of NIH-funded scientists over time, the results were startling. If trends continue, by 2020 there will be more investigators over 68 than under 38 (see p. 848). "If we don't fund the pipeline now, we will pay for it 20 years from now," Zerhouni says.

Zerhouni created special awards for young scientists but concluded that wasn't enough. In 2007, he set a target of funding 1500 new-investigator R01s, based on the previous 5 years' average. Some institutes struggled to reach their targets, NIH officials say. At the National Institute of Neurological Disorders and Stroke, for example, the shift to new grants meant that only 9% to 10% of established investigators with strong peer-review scores received funding, whereas 25% of comparable new investigators did, says NINDS Director Story Landis. She maintains, however, that "it's not as though a huge number of investigators lost out."

Some program directors grumbled at first, NIH officials say, but came on board when NIH noticed a change in behavior by peer reviewers. Told about the quotas, study sections began "punishing the young investigators with bad scores," says Zerhouni. That is, a previous slight gap in review scores for new grant applications from first-time and seasoned investigators widened in 2007 and 2008, Berg says. It revealed a bias against new investigators, Zerhouni says.

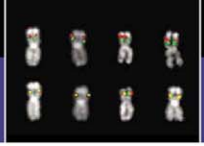
The 2007 target had an immediate effect: For the first time since 1995, new investigators and established ones submitting new grant applications had nearly the same success rate, about 19%. (Investigators renewing existing grants still do much better, however.) From now on, NIH will set award targets designed to equalize new grant success rates for the two groups.

NIH will also fine-tune its policy to tilt it in favor of early-career scientists. The goal is to adjust for the recently discovered fact that only about 55% of investigators who receive their first NIH grants are at an early stage of their career. The rest ▶



A leg up. After NIH set a numerical target for grants to first-time investigators in 2007, the number of awardees grew. Their success rates matched those of established investigators seeking new grants.

CREDITS: NIH



An intriguing stretch of DNA
842

Productive after all these years
848

are scientists who had been funded by other agencies or came from NIH's intramural program or from Europe after being forced to retire there. "It was embarrassing" to realize, for example, that the new investigators included two department chairs with Veterans Administration funding, Landis says. The targets will favor "early stage investigators," defined as researchers within 10 years of finishing their Ph.D. or residency.

Those outside NIH are generally supportive of the new-investigator targets, which were also endorsed earlier this year by an advisory committee reviewing NIH's peer-review policies (*Science*, 29 February, p. 1169). But at the same time, some scientists may be uneasy about the cost, says Howard Garrison, spokesperson for the Federation of American Societies for Experimental Biology in Bethesda, Maryland: "Every time you give a leg up to a young investigator, you're pushing someone off the edge of the cliff." Some observers say the real test will come when early stage investigators try to renew their grants: They may have trouble, and gains in creating a more youthful corps of investigators could be lost (*Science*, 26 September, p. 1776). NIH officials say they've looked at the data, and so far it seems that first-time investigators do just as well as established investigators who are renewing a new grant.

–JOCELYN KAISER

CLIMATE CHANGE

Rules for Ocean Fertilization Could Repel Companies

An international body has for the first time placed restrictions on experiments designed to fertilize large swaths of the world's oceans with a view to combating global warming. Meeting last week in London, delegates from 85 nations noted that such experiments "may offer a potential strategy for removing carbon dioxide from the atmosphere" by producing algal blooms that would absorb CO_2 and sink to the ocean floor. But they limited the experiments to "legitimate scientific research," a phrase not yet defined that could complicate plans to commercialize the approach.

Created in 1972 under the auspices of the United Nations' International Maritime Organization, the London Convention Treaty is supposed to regulate pollution in international waters. Members of the convention and the related London Protocol had been silent on ocean fertilization until several companies announced plans last year to carry out large-scale tests, setting off concern about environmental effects. The companies hope the technology will allow them to sell carbon credits on domestic and international markets.

On 31 October, delegates agreed unanimously to set scientific guidelines for proposed fertilization experiments, taking into account their expected carbon flux, impacts on oxygen levels and food webs, and the possibility that they will promote the growth of toxic species. Scientific bodies affiliated with the treaty will meet in May to hash out details.

"There was a widespread recognition among the delegates that there should not be a ban on legitimate research," says Henrik Enevoldsen, who observed the 5-day negotiations as a scientific staff member of UNESCO's International Oceanographic Commission. But the reference to "legitimate" studies was intended by some nations to exclude for-profit fertilization efforts, he says. "Most countries are looking to oppose something that's commercial research with an eye toward obtaining carbon credits."

Not ironed out. New global guidelines are being drawn up to govern experiments like this 2002 release of iron in the Southern Ocean.

CREDIT: MBARI

The experiments dump elements such as iron or nitrogen into the open ocean to stimulate the growth of plankton blooms (*Science*, 30 November 2007, p. 1368). Up to 3 tons of iron at a time have been released in a dozen small-scale fertilization experiments since 1993, and prominent scientists believe the technique, if scaled up, could sequester up to 1 billion tons of carbon dioxide per year as the blooms grow and die. But there are no international rules to regulate the practice, and researchers have identified myriad possible side effects, including local disruption of marine ecosystems or emissions of nitrous oxide, a potent greenhouse gas.

To quantify both the promise and perils of ocean fertilization, scientists want to launch experiments 10 to 30 times larger than earlier tests. Last week's vote implicitly supports such work, says geochemist Ken Buesseler of Woods Hole Oceanographic Institute in Massachusetts. But Buesseler worries that the upcoming rules, which will form the basis for permits to be issued by individual countries, "could preclude even legitimate science, if the [environmental] assessment needs to include measurement of all impacts on all time [and] space scales."

On the other hand, Greenpeace and other environmental activist groups are concerned about possible bias in reporting results among commercial companies looking to fight global warming by exploiting ocean fertilization for profit. But Dan Whaley, CEO of Climos, a San Francisco, California–based ocean fertilization start-up, defends his company's ethics and notes that the text of the resolution doesn't explicitly bar commercial projects. Climos still hopes to abide by the treaties and obtain permits for operations previously scheduled in 2010.

–ELI KINTISCH

CLIMATE CHANGE

Chinese Cave Speaks of a Fickle Sun Bringing Down Ancient Dynasties

A 1.2-meter-long chunk of stalagmite from a cave in northern China recorded the waning of Asian monsoon rains that helped bring down the Tang dynasty in 907 C.E., researchers report on page 940. A possible culprit, they conclude: a temporary weakening of the sun, which also seems to have contributed to the collapse of Maya civilization in Mesoamerica and the advance of glaciers in the Alps. "I think it's one of the coolest papers I've seen in a long time," says paleoclimatologist Gerald Haug of the Swiss Federal Institute of Technology in Zurich. This latest cave record also points to the potentially devastating effects that climate change—even change that's mild when averaged around the globe—can have on vulnerable local populations.

Although hardly the final word in such controversial fields, the cave record—which other researchers describe as "amazing," "fabulous," and "phenomenal"—provides the strongest evidence yet for a link among sun, climate, and culture. The key to obtaining it was "a really, really clean sample," says paleoclimatologist Lawrence Edwards of the University of Minnesota (UM), Twin Cities. Paleoclimatologists Pingzhong Zhang of Lanzhou University in China, Hai Cheng of UM, Edwards, and colleagues collected a stalagmite (a mound composed mostly of calcium carbonate slowly precipitated from dripping groundwater) from Wanxiang Cave in northern China at the far reach of the rains of the summer Asian monsoon.

Relatively high amounts of uranium and exceptionally low clay-borne thorium in this stalagmite enabled them to conduct uranium-thorium radiometric dating of the layered deposits to within an average of just 2.5 years. As a result, they could calculate precise dates for subtle variations in the stalagmite's oxygen isotope composition that reflect variations in rainfall near the cave. "They absolutely nailed the rainfall history of [northern] China over the past 1800 years," says Haug.

Comparing their rain record with Chinese historical records, Zhang and colleagues found that three of the five multicentury dynasties during that time—the Tang, the Yuan, and the Ming—ended after several decades of abruptly weaker and drier summer monsoons, possibly poor rice harvests, and social turmoil. In turn, decades that included the strongest, wettest monsoon of the past millennium coincided with the Northern Song Dynasty's golden age of rich harvests, exploding population, and social stability. "Our results really match the historical record," says Edwards. "You can't figure it's all climate, but when you see these nice correlations, you see that climate probably played an important role."

The group then looked farther afield. Critical parts of their monsoon rainfall record—in particular the dryness of the late Tang dynasty—match neatly with a previously published climate record from a lake on the southern coast of China, with the advances and retreats of Swiss alpine glaciers, and with records from within and near Central America. Most striking is the correlation between the Asian monsoon and the collapse at the end of the Maya Classic period under severe drought duress around 900 C.E. (*Science*, 18 May 2001, p. 1293), near the end of the drought-stricken Tang dynasty.

Previous research had linked changes in ▶

Good times. Monsoon rains were plentiful early in the Northern Song Dynasty of China, according to the isotopic record in a cave stalagmite (*top*). A cave-wall painting from the same province (*above*) recorded the bounty.

CREDITS: (TOP TO BOTTOM): COURTESY OF HAI CHENG; THE BRIDGEMAN ART LIBRARY

Call to Resume Nutrition Program

PARIS—After months of quiet diplomacy, Médécins Sans Frontières (MSF) has issued a public call to President Tandja Mamadou of Niger to let the humanitarian organization resume its nutrition programs in the country. The suspension by the Nigerien government "endangers the lives of thousands of children," MSF said in a statement last week. The French section of MSF operated a massive program in Niger's central region of Maradi, where malnourished children were given new, peanut-based products said to have revolutionized malnutrition treatment (*Science*, 3 October, p. 36). But the Nigerien government ended the program in mid-July, accusing MSF of breaking rules for nongovernmental organizations and insufficient coordination with the national health care system. Negotiations have been fruitless.

The suspension has also interrupted research into the efficacy of the peanut pastes and a large-scale study of infectious diseases in malnourished children for which subjects were recruited from an MSF hospital. Scientists are hoping they can resume the latter study by recruiting patients from a local hospital instead, says MSF's Philippe Guérin.

–MARTIN ENSERINK

Stalking Killers in Africa

BEIJING—Virus hunters in West Africa are banding together to better cope with emerging threats and old foes. At the International Consortium on Anti-Virals (ICAV) meeting here this week, researchers from Nigeria, Ghana, and other countries in the region agreed to establish a West African Viral Surveillance Network. "This is a neglected area of the world," says ICAV co-founder Jeremy Carver, a molecular biologist and professor emeritus at the University of Toronto in Canada.

Organizers have not decided on goals for fundraising, which has just begun, so the focus for now is on forging connections. "Scientists in the region weren't talking with each other," says ICAV Africa director Oyekanmi Nashiru of the National Biotechnology Development Agency in Abuja, Nigeria. There is plenty to share. Nigeria has virology expertise but poor infrastructure, Nashiru says, whereas nearby Ghana has top-notch labs at the Noguchi Memorial Institute for Medical Research at the University of Ghana in Legon. Key quarry include bird flu, HIV, and polio, which has yet to be eradicated from West Africa. When it comes to emerging viruses, says Noguchi Institute molecular biologist James Brandful, "now we'll be better prepared."

–RICHARD STONE

both the Asian monsoon and Mesoamerican climate to variations in the brightness of the sun (*Science*, 6 May 2005, p. 787). Checking their record, the group found an 11-year cycle in rainfall—the length of the shortest cycle of solar variability. In their record, rain tracked centuries-long trends in solar activity as measured in records of carbon and beryllium isotopes. And a climate model driven in part by solar variations broadly tracked the monsoon trends. "Solar variation is a player, but the sun is not everything," Edwards concludes. Internal jostlings of the climate system must also play a role, he says.

Climate modeler David Rind of NASA's Goddard Institute for Space Studies in New York City agrees. In a modeling study in press in the *Journal of Geophysical Research*, Rind and colleagues found that "the solar influence on the monsoon was more like a 'weighting of the dice'—it influenced the net result, but did not dominate," he writes in an e-mail.

De'er Zhang, chief scientist of the National Climate Center in Beijing, stresses that both climate and culture are too complex to be reduced to a simple cause-and-effect relationship. A single spot cannot properly represent such a vast area as encompassed by the monsoon, she writes in an e-mail, and numerous political factors influenced the Tang dynasty. "Climate might have played a role," she writes, but it was "far from playing 'a key role' as stated by [Pingzhong] Zhang *et al.*" Sorting out what "key" or "important" meant a millennium ago could require a lot more spelunking. **–RICHARD A. KERR**

PERSONAL GENOMICS

Number of Sequenced Human Genomes Doubles

Less than a decade ago, it took hundreds of millions of dollars and a large international community to sequence a single human genome. This week, three reports in the 6 November issue of *Nature* describe three more human genomes—the first African, the first Asian, and the first cancer patient to have their entire DNA deciphered. The sequences provide clues about genome variation and disease; they also demonstrate the potential of a relatively new sequencing technique to mass-produce human genomes. "The methods are extremely powerful," says geneticist James Lupski of Baylor College of Medicine in Houston, Texas. "Reading these papers, I think the personal genomes field is moving even faster than I anticipated."

Until now, four human genomes have been published: the reference human genome, derived from sequencing DNA from several anonymous individuals; one by Celera Genomics; and those of genome stars J. Craig Venter and James Watson. Efforts to date to identify differences among individuals have relied not on entire genome sequences but on surveys of single-base changes called SNPs and of structural variations in duplicated pieces of DNA (*Science*, 21 December 2007, p. 1842).

Even the broadest SNP surveys look at just a few million SNPs out of the 3 billion bases in the genome, leaving researchers in the dark about how much individual variation there is and how specific differences correlate with disease risks. Hence the push to drive down the cost of sequencing to $1000 per genome (*Science*, 17 March 2006, p. 1544). The newly published genomes came in with price tags of $250,000 to $500,000 each but would cost half that or less if done today. The three groups all used a technology developed by Solexa, now part of Illumina Inc. in San Diego, California, to speed and slash the cost of sequencing. It generates smaller pieces of sequence faster and cheaper than previous technologies. Such small pieces used to be difficult to stitch together, but this approach can work well now because the reference genome helps guide their assembly.

New genome on the block. The first genome sequence from a Chinese was on display last year at a technology fair in Shenzhen, China.

CREDIT: XINHUA/LANDOV

To explore the genetic underpinnings of cancer, Richard Wilson and colleagues at the Washington University School of Medicine in St. Louis, Missouri, sequenced genomes from both normal skin tissue and tumor tissue of a middle-aged woman who died of acute myelogenous leukemia (AML). They compared the DNA to determine what was different about the cancer cells. About 97% of the 2.65 million SNPs found in the tumor cells also existed in the normal skin cell, suggesting they were not critical to the cancer process. The researchers also eliminated SNPs that had been previously identified elsewhere as well as those that did not change the coding of a gene, ending up with 10 SNPs unique to the tumor cells. "I don't think we missed anything," says Wilson.

Two occurred in genes previously linked to this leukemia. Eight led the researchers to new candidate AML genes, including several tumor suppressor genes and genes possibly linked to cell immortality. By sequencing the whole cancer genome, "we capture what we don't know as well as what we do know [about cancer genes]," says Illumina's David Bentley. "That can really transform our ability to understand cancer."

Bentley and colleagues sequenced the genome of a Yoruba man from Nigeria whose DNA has already been extensively studied, enabling them to check the accuracy of their technology. In the third *Nature* paper, Jiang Wang of the Beijing Genomics Institute in Shenzhen, China, and colleagues sequenced the genome of a Han Chinese male. The Yoruba analysis uncovered almost 4 million SNPs, including 1 million novel ones. The Chinese genome had about 3 million, including 417,000 novel SNPs. As anticipated, the African genome had greater variation per kilobase than either the Chinese or sequenced Caucasian genomes, indicative of its ancestral status.

These new genomes were already significantly cheaper than their predecessors were; next year, Illumina expects the cost to drop to about $10,000. Other companies are promising even lower prices per genome. Nonetheless, geneticist Aravinda Chakravarti of Johns Hopkins University School of Medicine in Baltimore, Maryland, is cautious about how quickly genome sequencing should enter the clinic: "We still don't know how to interpret [the data]," he notes. Bentley agrees. Because of the uncertain applicability and utility of sequence data, "and possibly ethical barriers," he notes, saying the technology is poised to enter the clinic anytime soon is "pushing it." **–ELIZABETH PENNISI**

PERSONAL GENOMICS

The Touchy Subject of 'Race'

Nothing makes scientists more nervous than the topic of "race," so much so that they'd like to find a way not to talk about it at all. That was the core issue last week at a meeting* at the National Human Genome Research Institute (NHGRI) in Rockville, Maryland, where about 40 scientists and ethicists debated how to present the torrent of new findings from human gene sequencing studies to the public.

In different parts of the world, different gene mutations become advantageous and spread quickly through a population, making some variants more prevalent in particular ancestral groups. Some are innocuous enough—such as the emergence of lactose tolerance in farming populations. But there's already much debate over the use in medicine of findings of racial differences in the prevalence of genes associated with certain diseases. Many scientists predict that it won't be long before they have solid leads on much more controversial genes: genes that influence behavior—possibly including intelligence.

Everyone at the meeting agreed on the need for non-"fraught" terminology—"geographic ancestry," for example, instead of "race." But specifying such ancestries is also a minefield. "Amerindian," for example, is offensive to Native Americans, according to one speaker. "Caucasian" is also unacceptable because it implies racial rather than geographic ancestry. Some speakers even advised that it is inappropriate to refer to a "European allele" for lactose tolerance, because it also occurs in other groups.

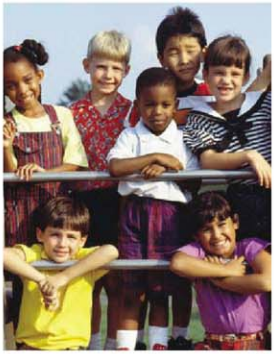
Participants acknowledged that however they characterize their findings, they can't control what the public makes of them. "When translated into popular culture, society reads whatever term we pick as 'race,'" said Timothy Caulfield, a health law professor at the University of Alberta in Edmonton, Canada. Carlos Bustamante, a population geneticist at Cornell University, said that when his group published a study in *Nature* this year indicating that European-Americans had more deleterious gene mutations than African-Americans, some publications touted the report as suggesting that blacks are fitter than whites.

Some tense moments came during a discussion of a paper on brain genes. In 2005, geneticist Bruce Lahn and colleagues at the University of Chicago in Illinois reported evidence for selection in mutations of two genes regulating brain development that are more common in Eurasians than in Africans (*Science*, 9 September 2005, pp. 1717 and 1720). They hypothesized that these mutations were related to the human cultural explosion some 40,000 years ago (*Science*, 22 December 2006, p. 1871). Celeste Condit, a professor of speech communication at the University of Georgia, Athens, criticized the way the papers were written, saying they could be seen as having a "political message embedded" in them: that the genes might contribute to racial differences in brain size and therefore perhaps to racial differences in IQ. Lahn denied any political message, telling her she was "putting words in [my] mouth."

Later, Lahn commented that some scientists "are almost like creationists" in their unwillingness to acknowledge that the brain is not exempt from selection pressures.

At the end of the day, Allen Buchanan, a philosophy professor at Duke University in Durham, North Carolina, warned the group against going overboard. "A visible, concerted effort to change vocabulary for moral reasons is likely to trigger a backlash," he said. There's "risk of … stifling freedom of expression in the name of political correctness," he said, and losing credibility in the process.

–CONSTANCE HOLDEN



CREDIT: PHOTOS.COM

Ancestry, not race. Researchers are grappling with how to communicate genetic data on differences among populations.

*Workshop on Ethical, Legal, and Social Issues in Natural Selection Research.

A Graduate Appetizer

The National Academies' eagerly awaited assessment of U.S. doctoral programs won't be released for another few months. But for the university administrators, faculty members, and graduate students whose lives are influenced by this mammoth undertaking, a description of what's new since the 1995 edition should be available in a few weeks. One wrinkle will be multiple ratings: In addition to a score from peers on overall quality, each of the 5000-plus programs at 212 universities will be ranked according to dimensions such as faculty productivity, diversity, and student outcomes. "They include factors the schools can influence and those that they can't really control," says study director Charlotte Kuh of the National Research Council. –JEFFREY MERVIS

Backing Up Hubble

The good news for NASA's Hubble Space Telescope is that controllers last week finally got a balky backup system to take over the job of sending images to Earth after the main system malfunctioned. The bad news is that managers have tacked on several months to the scheduled launch of a mission intended, among other things, to replace the faulty data system and avoid dependence on the backup. In a 31 October press conference, NASA officials said that preparing a replacement data system for launch and installation by astronauts will delay the repair mission, Hubble's last upgrade, until at least May 2009. –ANDREW LAWLER

A Basic Change for Korea

South Korea's academic researchers are smiling in anticipation of next year's budget. The Ministry of Education, Science, and Technology is seeking a 9.5% rise in its 2009 budget to $3.2 billion. "Korea has concentrated on applied science until now, but governmental policy is changing to increase support for basic research" to produce fundamental breakthroughs for technological development, explains Hang Sik Park, director of the ministry's Science and Technology Policy Planning Bureau. He says the proportion of governmental funding going to fundamental research could rise to roughly 28% of the total, up from about 25.6% this year. Other areas in line for big funding boosts include international collaborations, rising 75% to $34 million, and green technologies, with a 92% increase to $53 million. Other ministries have not yet announced their R&D requests, but the budget will go before the national assembly in December.

–DENNIS NORMILE

FOUNDATIONS

Economic Woes Threaten to Deflate Plans for 2009

Uncertainty has become the new norm for economic forecasters. But scientists planning next year's experiments want to know how the stock market turmoil, a credit crunch, and a recession will affect their research. It's an urgent question, especially with the U.S. government facing a yawning deficit and a likely squeeze on domestic spending. Among the first to feel the slowdown are charitable foundations and other philanthropies, which provide billions of dollars in funding to scientists each year, including support for innovative, risky research that the government may be reluctant to back. Some are scaling back; some say they're holding steady. Others say they cannot plan far ahead—not even to predict what the next 2 months, normally flush fundraising time, will bring.

"I've been in this business 30 years, and I've never seen an environment" like this, says Richard Mattingly, executive vice president and chief operating officer of the Cystic Fibrosis Foundation, which in 2008 gave out $199 million in research money. The foundation relies exclusively on fundraising. Mattingly, emerging from a board meeting last week, said he expects funding to drop next year, though he can't yet say how much.

One of the hardest-hit organizations so far is the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation in Needham, Massachusetts, which has delayed $65 million in research funding to dozens of investigators for the second half of 2008 and 2009. Projects in melanoma, lymphoma, neurodegenerative diseases, inflammatory bowel disease, and others were ready to go, says Bruce Dobkin, the group's executive director and a neurologist at the University of California, Los Angeles. "Now, we're going to have to wait and see what happens." Dobkin says he doesn't know what, precisely, prompted such drastic action, and the foundation declined to comment further.

The Nature Conservancy, with more than 600 scientists on staff, decided 3 weeks ago to cut its current research budget by 10% and laid off some scientists. Peter Kareiva, the group's chief scientist, says international programs will be especially hard hit. The group is drawing up contingency plans in case further cuts are necessary.

Groups that rely mainly on contributions are nervously entering their peak fundraising season. Last week, the Multiple Myeloma Research Foundation (MMRF) in Norwalk, Connecticut, held its biggest gala of the year, at the Hyatt Regency Greenwich Hotel in Connecticut, with supermodel Cindy Crawford. Nine hundred people accepted; the event was expected to raise $1 million—impressive, but below the $1.5 million originally hoped for, says Scott Santarella, the group's chief operating officer. MMRF scaled back its research funding several months ago for 2008, cutting it from $17 million to $15 million, but expects to bring it back up to $17 million in 2009. The group, like many charities, typically raises 40% of its money in the last quarter of the year.

Similarly, the Michael J. Fox Foundation for Parkinson's Research held its star-studded gala this week in economically battered New York City, with the English rock band The Who performing. Leaders hope to raise a bit over $4 million from that event, down from last year's $5 million, says co-founder Debi Brooks.

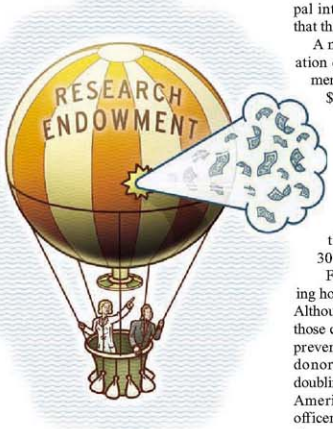


ILLUSTRATION: PETER HOEY

For organizations that live off endowment income, the drop in their value can be dizzying: The Burroughs Wellcome Fund fell from nearly $700 million at the end of July to $540 million last Friday, and the Bill and Melinda Gates Foundation lost $3.6 billion between the beginning of this year and the end of September, to end with $35.1 billion. The Howard Hughes Medical Institute's (HHMI's) worth fell from $18.7 billion in August 2007 to $17.4 billion at the end of August this year—and that was before the big drop in the stock market. (Neither the Gates Foundation nor HHMI would release current figures.) "Sometimes you go into a meeting and by the time you come out the endowment's gained $10 million, and by the end of the day it's lost $20 million," says Burroughs Wellcome spokesperson Russ Campbell, describing the wild gyrations in the market.

HHMI is required by law to distribute 3.5% of its assets each year, and foundations like Burroughs Wellcome must give out 5%. This is normally not a problem because these groups offset the outflow with investment gains, keeping their principal intact. Not all may be able to manage that this year.

A mid-September survey by the Association of Small Foundations (ASF), whose members have an average endowment of $20 million and give away $1 million each year, found that 84% said their endowments had dropped this year. But, responding days after the investment bank Lehman Brothers collapsed, 64% said they plan to maintain or increase grant budgets in 2009. That said, "I do get e-mails that say, 'Oh my gosh, we're down 30%,'" admits Tim Walter, ASF's CEO.

Fundraisers, meanwhile, are considering how to persuade donors to keep giving. Although many will continue to send checks, those checks may be smaller than before. To prevent that, and to stem the departure of donors altogether, many groups are redoubling their communication efforts. At the American Cancer Society, chief medical officer and oncologist Otis Brawley recently disseminated a list of 10 scientific discoveries funded with ACS dollars. "We have not made any plans right now to decrease our funding for cancer research," says Brawley, and in fact he says research funding is up about 5% this fiscal year, which began in September, over last. But although he doesn't work closely with ACS fundraisers, "I see those guys on the elevator, and they're not happy."

—JENNIFER COUZIN

With reporting by Jon Cohen.

17q21.31: Not Your Average Genomic Address

This one region of chromosome 17 has had a storied history, with changes in its DNA of import to human evolution and disease

For most of us, 17q21.31 is a meaningless alphanumeric. For geneticists, it's a genomic postal code identifying a region of chromosome 17. But for Tjitske Dansen, a Dutch mother of three, it's an answer for which she waited 17 years. From birth, her oldest daughter, Anne Zandee, had trouble. "She kept lagging in many respects: walking, talking, growing," Dansen recalls. For a while, the toddler had epileptic seizures. Weak jaw muscles cause Zandee to drool; weak back muscles may have contributed to her scoliosis. "But we never knew what was wrong with her."

Four years ago, an orthopedics doctor referred Zandee to Bert de Vries, a clinical geneticist at Radboud University Nijmegen Medical Centre (RUNMC) in the Netherlands, to try to find a genetic explanation for the scoliosis. "There had been so many studies, and they never found anything, so we didn't think anything would come of it," says Dansen. "We didn't hear anything from De Vries for a year and a half, when suddenly he called us." He told them Zandee was missing a piece of chromosome 17; to be exact, a piece of 17q21.31. Zandee is now one of 22 documented cases of a new genomic disorder. "We were happy to find out," says Dansen. But, "of course we had never heard of 17q21.31 before."

Among geneticists, however, 17q21.31 has been gaining notoriety for almost 20 years. Its half-dozen genes include one controversially implicated in Alzheimer's disease and firmly tied to other dementias. More recently, researchers excavating this single chromosomal address, located about 19 million bases down the "q," or longer arm, of chromosome 17, have uncovered a tumultuous past. Here, vulnerable DNA has gone astray to cause mental retardation, learning disabilities, and even cancer. Genes hide within genes, and variation in this sequence even suggests to a few researchers that our species interbred with Neandertals. "It's probably one of the most bizarre and fascinating regions of the human genome," says Evan Eichler, a geneticist at the University of Washington, Seattle.

Zoom, zoom

The most famous gene that lives at this address is *MAPT* (*microtubule-associated protein tau*). Tau first drew neuroscientists here because it is the protein that gets jumbled together to form

DNA locator. Under a microscope, gene-rich and gene-poor regions stain differently, creating readily defined and labeled genomic addresses. 17q21.31 is in red.

CREDITS: COURTESY OF ROB ZANDEE; SOURCE: GENE CARDS

Short on DNA. From birth, Anne Zandee's development lagged. A deleted piece of chromosome 17 is to blame.

neurofibrillary tangles in the brains of patients with Alzheimer's disease. But even though Athena Andreadis, now at the University of Massachusetts Medical School's Eunice Kennedy Shriver Center in Waltham, and her colleagues cloned *MAPT* in humans in 1992, neither they nor others have been able to find mutations that could explain Alzheimer's. Most considered further investigations a waste of time and lost interest in 17q21.31.

> *"It's probably one of the most bizarre and fascinating regions of the human genome."*
>
> —EVAN EICHLER, UNIVERSITY OF WASHINGTON, SEATTLE

But John van Swieten of Erasmus University Medical Center in Rotterdam, the Netherlands, was eyeing that gene region with another disorder in mind: Pick's disease, a neurodegenerative condition in which individuals become increasingly bored, listless, and incapable of relating to others. Personal hygiene fails, emotions falter, and symptoms become progressively worse until full-time care and supervision are required. In this disease, also called frontotemporal dementia (FTD), the frontal lobe of the brain shrinks and tangles form.

In 1994, Kirk Wilhelmsen and Timothy Lynch of Columbia-Presbyterian Medical Center in New York City established a link between FTD and 17q21.31 by evaluating how the disease was inherited in one large family. Four years later, Van Swieten, Erasmus geneticist Peter Heutink, and Michael Hutton, now at Merck Research Laboratories in Boston, pinpointed mutations in *MAPT* responsible for 10% of the cases of this disease.

Suddenly, the gene had sex appeal. "It provided a rationale for why tau was important in Alzheimer's," and it became possible to develop mouse models to study tau's effects, recalls Hutton. As a result, "a lot of people moved to the field," says Van Swieten.

Hutton, who now devotes his career to looking for potential Alzheimer's disease therapies that target tau tangles, was investigating yet another tau-related disease. In 1999, while sequencing the *MAPT* gene from patients with progressive supranuclear palsy, he and his colleagues noticed something odd. "We got interested in it almost as a piece of DNA rather than its relationship with the tangles," recalls Hutton's collaborator, John Hardy of University College London.

Most chromosomes undergo recombination: During cell division, bits of one chromosome swap places with comparable bits of the matching chromosome, introducing small differences in the DNA sequence from one generation to the next and between one individual and another. The pattern of those differences is called the haplotype. What struck Hutton, Hardy, and their colleagues was that there seemed to be two very distinct haplotypes in the *MAPT* region. One, dubbed H1, seemed to be slightly variable, indicative of some recombination. But the sequence of the other, H2, was nearly identical across about 1.3 million bases in everyone with that haplotype, at least at all of the bases they examined. "It was inherited as one long lump of DNA," Hardy explains.

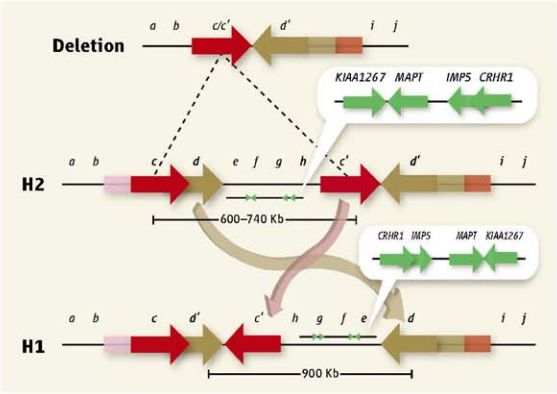
Hutton and Hardy realized there was something very odd about H2. "There were clearly unusual structures in or close to the boundary of the haplotype block," Hutton recalls. Researchers in Iceland were coming to a similar realization, and eventually they scooped the British group in making a startling determination: Almost a million bases in H2 were pointed in the wrong direction.

Kári Stefánsson and his colleagues at deCODE Genetics in Reykjavik, Iceland, had also noticed the lack of variability along 17q21.31 in some individuals. When Stefánsson's group took a close look at the reference human genome sequence at this location, they realized that the sequence, which had involved deciphering DNA from multiple individuals to come up with a consensus genome, contained bits of both H1 and H2. So, they went back to the drawing board to sort out the differences between the two.

A comparison of separate H1 and H2 sequences revealed that H2 has a 900,000-base stretch of 17q21.31 that is inverted relative to H1, the deCODE group reported in 2005. The boundaries, or breakpoints, of the inverted region consist of low-copy repeats, blocks of DNA duplicated multiple times. Based on a comparison with chimpanzee DNA from the same region, the researchers concluded that the second haplotype emerged at least 2 million years ago.

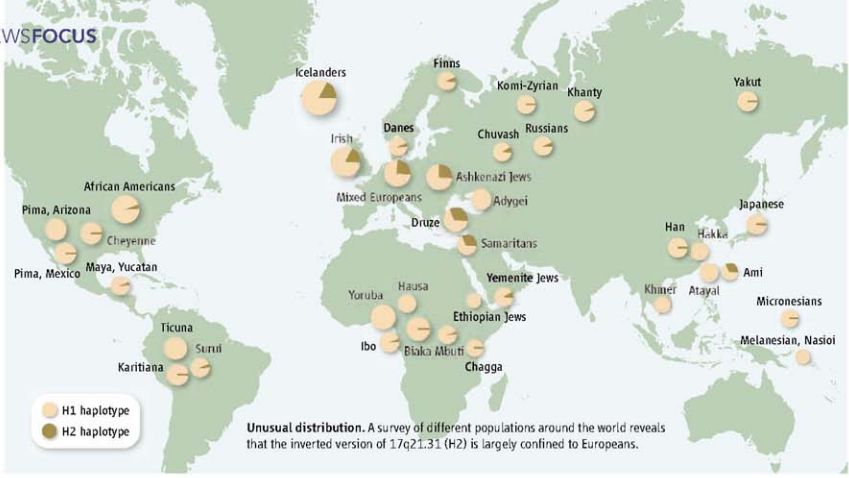
The haplotypes were not evenly distributed, however. Most people are H1. Stefánsson and, independently, Hardy's group found H2 almost exclusively in Europeans, at a frequency of about one in five. "Since the inversion is largely restricted to Caucasians, we all thought the ancestral state would be the H1 orientation," says Eichler. Indeed, the deCODE data suggest that once the inversion occurred, H2 spread because it provided a reproductive edge, says Stefánsson: Women with H2 had more children than women with H1, they reported in 2005. They saw a similar, but less clear-cut, trend for men. These results implied that H2 was under positive selection and should be on the rise.

There was a problem with this scenario, though: Some data indicated that H2 predated H1. In August, Eichler and his colleagues showed that was indeed the case. Eichler's group sequenced both H1 and H2 and carried out a detailed comparison among



Flip and fall. The H2 version of this genomic region has genes (green) and duplicated regions (thick arrows) facing the wrong way compared to H1. This orientation predisposes H2 to losing a gene-rich section of DNA (Deletion).

CREDIT: ADAPTED FROM LUPSKI, *NATURE GENETICS* 38, 1141 (2006)



Unusual distribution. A survey of different populations around the world reveals that the inverted version of 17q21.31 (H2) is largely confined to Europeans.

the two human versions of the region and the same stretch of DNA in chimpanzees, macaques, and orangutans. All three macaque species they examined and the Sumatran orangutan carried only the inverted version. The two chimp species carried a mix of inverted and non-inverted copies, with the inverted version predominating, and the Bornean orangutan has both as well. H2 is the more ancient haplotype, Eichler and his colleagues concluded in a paper published online by *Nature Genetics* on 10 August. "It's an amazing result," says Hutton.

The sequence comparisons also reveal that independently in humans, chimps, and orangutans, this 900,000-base region has reoriented itself into the H1 orientation, which explains why Eichler found both orientations in these primates. "This bit of DNA has been flip-flopping up and down. There must be an evolutionary reason for that, but we don't know what it is," says Hardy.

Eichler suspects that when H1 appeared, it somehow provided a strong fitness bonus and became much more common over time at the expense of H2. In Africans, H2 almost disappeared, except in the relatively few people who migrated to Europe 50,000 to 100,000 years ago. Then, for as-yet-unknown reasons, H2 provided its own advantage in the European population—as Stefánsson's data show—and the pendulum has begun to swing in the other direction.

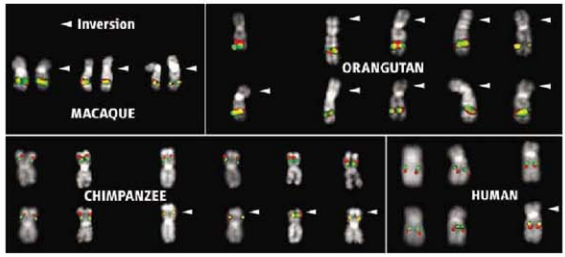
Hardy and, to a lesser extent, Stefánsson give credence to a more extreme explanation for the distribution of H2. Hardy thinks that H2 had disappeared from the modern humans moving out of Africa to populate the Northern Hemisphere but not from Neandertals, who reintroduced the inversion into the European gene pool through interbreeding with *Homo sapiens* 28,000 to 40,000 years ago. This view is not supported by the genetic evidence emerging from sequencing Neandertal DNA, and "I realize it's an off-the-wall idea," says Hardy. But he nonetheless thinks it's plausible.

Whatever happened, about 50,000 years ago, H2 went haywire, with duplicated regions begetting ever more duplicated regions. Low-copy repeats can destabilize a chromosome by confusing the DNA recombination machinery and causing the repeat regions to be copied extra times. Repeats can also cause skips, resulting in DNA between repeats getting left out. This had happened in H2 but not as much in H1. H2 has 441,000 bases' worth of repeats at the boundaries of the inverted DNA, compared with 169,000 bases in H1. H2 also carries extra duplications within its boundaries and is organized in such a way as to predispose the sequence to further rearrangement.

Lost DNA

As Dansen and her daughter Zandee know all too well, those extra duplications can spell trouble. They were the tip-off, not just for De Vries but also for other researchers trying to understand mental retardation, that drew them to the microdeletion responsible for Zandee's syndrome.

In 2002, Eichler and his colleagues were cruising the genome in search of repeats, or segmental duplications—nearly identical stretches of genome at least 10,000 bases long, each separated by 50,000 to 10 million bases—thinking they might mark places where the genome was in disarray and causing



Which came first? Readily distinguishable red and green tags merge and look yellow in chromosome 17 containing inverted DNA. Labeled macaque, chimp, orangutan, and human chromosomes reveal that the inversion dates deeper in the primate tree than the noninverted version.

CREDITS: ADAPTED FROM HREINN STEFANSSON, ET AL., *NATURE GENETICS* 37, 129 (2005); MICHAEL ZODY ET AL., *NATURE GENETICS* ADVANCE ONLINE PUBLICATION; DOI 10.1038/NG.193

disease. They came up with 130 possible problematic spots, then investigated those spots in 290 people with mental retardation. They found 16 rearrangements, including four people with a piece of 17q21.31 missing. Using microarrays, they figured out that when the missing DNA dropped out, it took a half-dozen genes with it, including *MAPT*. The breakpoints are two 38,000-base low-copy repeats flanking the DNA deleted in these individuals.

Parents have the inversion, which is necessary to set up the repeats in such a way that deletions become more likely. "Some think it's the inversion itself that's the culprit, but it's not," says Eichler. It's the large number of repeats and their orientation that make 17q21.31 vulnerable.

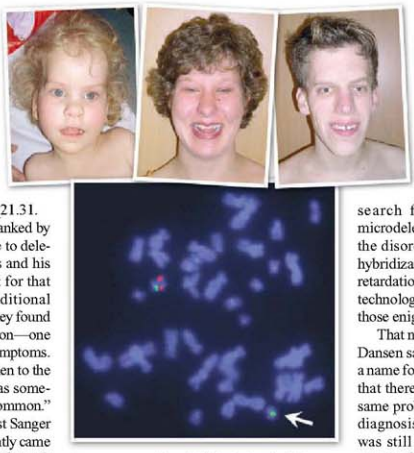
De Vries came to this microdeletion by a different route. Eager to help parents understand the basis of unexplained mental retardation in their children, De Vries and his colleagues initially screened 340 patients using a technique called microarray-based comparative genomic hybridization to detect genomic rearrangements too small to see by simply staining the chromosomes. In one, they detected a missing piece of 17q21.31.

Because the missing piece was flanked by low-copy repeats that might give rise to deletions in other individuals, De Vries and his colleagues designed a probe to test for that missing piece and screened an additional 840 mentally retarded individuals. They found two more people with the same deletion—one was Zandee—and the same set of symptoms. "We went first to the genotype and then to the phenotype," says De Vries. "This was something new, but it will become more common."

Nigel Carter of the Wellcome Trust Sanger Institute in Hinxton, U.K., independently came across this microdeletion syndrome through queries to a database called DECIPHER. Entries include comparative genomic hybridization results, along with clinical descriptions of the symptoms of the people tested. "Many of us as clinicians may see one of these kids in our lives, but [these researchers] got the same descriptions with the same array findings from three [entries]," says James Lupski, a geneticist at Baylor College of Medicine in Houston, Texas.

The three teams published independent reports back to back in 2006 in *Nature Genetics*. Now they have joined forces to describe 22 patients in molecular and clinical detail in a paper published online 15 July by the *Journal of Medical Genetics*. They calculate the prevalence of this new genomic disorder to be 1 in 16,000 newborns, and it may account for up to 0.64% of unexplained mental retardation in Europeans. "This is the first novel microdeletion syndrome identified and one of the most frequent ones," says collaborator Joris Veltman, a molecular geneticist at RUNMC.

The deleted region contains six genes, and at this point, they don't know which loss matters the most. Even so, "we've gone from 2 years ago not even knowing the syndrome existed to having [dozens of] kids [diagnosed]," says Lupski. "We're going to see that more and more and more." Last year, Danish clinical geneticists came across a patient with unexplained mental retardation whose abnormality was an extra copy of what was deleted in De Vries's patients.



Common cause. These individuals share facial features indicative of a DNA deletion, visualized by the absence of a red tag in a patient's stained chromosomes.

CREDITS: DAVID KOOLEN *ET AL.*, *NATURE GENETICS* **38**, 999 (2006); SHARP *ET AL.*, *NATURE GENETICS* **38**, 1038 (2006)

Still a puzzler

Despite this quick success, 17q21.31 is still slow to give up its secrets. It is clear that both haplotypes have their pitfalls: H1 increases the risk of progressive supranuclear palsy and other neurodegenerative diseases, likely by increasing the production of *MAPT*, and H2 increases the chances that offspring will have mental retardation because of a microdeletion. But at every turn, this genomic address proves a little more complicated.

Consider the elusive Alzheimer's connection, where no causative *MAPT* mutations have yet been found. Frustrated, Christopher Conrad, a neurogeneticist at Columbia University, began looking for other undiscovered genes in that region. In 2001, he found a very tiny one inside *MAPT* that bears no resemblance to any known gene. "It's one of the few examples of a gene within a gene," says Conrad. He named it *Saitohin*, after the deceased adviser who helped get him started on the project, and has spent the past several years trying to figure out its role. The gene seems to have appeared first in primates, and Andreadis, who is collaborating with Conrad, has determined that it interacts with a protein involved in antioxidation. It seems to lead to alternative splicing of *MAPT*, which may result in a version of tau that is more likely to aggregate into tangles. "Given the appearance only in primates, it's tempting to say the gene could have something to do with brain development," says Conrad.

Likewise, Stefánsson is frustrated by 17q21.31's enigmatic connection to psychiatric disorders. "We've done a lot of work to see what the risk [of the inversion] is to schizophrenia, but we have not succeeded yet," says Stefánsson.

De Vries is continuing to search for more individuals with the microdeletion syndrome and to characterize the disorder. Before comparative genomic hybridization, about half the cases of mental retardation went unexplained. Now, this new technology is making sense of about 10% of those enigmatic cases.

That makes a big difference to the parents. Dansen says her family was just glad to have a name for their daughter's disorder and to see that there were others just like her with the same problems. For at least one mother, the diagnosis brought good news. Her toddler was still not walking, but De Vries could reassure her that the others had also been slow to walk but did so eventually. "When I explained that to the mother, she was very relieved," says De Vries.

Now almost 20, Zandee plays in a special band, works in a canteen, and paints. Eventually, she will move from her parents' home to a group house with others with mental handicaps. "I will keep following the research," says her mother, although she's not sure what more it will tell her. But she knows that by tracking Zandee's progress, De Vries will learn a lot about adult diseases associated with the syndrome and even about life expectancy. "My son recently asked, 'How long can she live anyway?'" says Dansen. "We have no idea. Nobody does." **–ELIZABETH PENNISI**

With reporting by Martin Enserink.

NEUROTECHNOLOGY

Engineering a Fix for Broken Nervous Systems

A recent meeting on neural prosthetics provided an update on progress and some interesting digressions

PALO ALTO, CALIFORNIA— "I believe we're at the beginning of a new age of neurotechnology," Brown University neuroscientist John Donoghue told researchers who gathered here recently to discuss the state of the art in neural prosthetics, surgically implanted devices designed to restore sight to the blind, hearing to the deaf, and movement to paralyzed people. The idea of engineering a fix for nervous systems that can't heal themselves continues to spur both hope and hype; the meeting here at Bio-X, Stanford University's interdisciplinary research center, provided a glimpse of where the technology really stands. It also prompted frank discussions of current challenges and some fascinating, if slightly tangential, dinner conversations.

If the age of neurotech is indeed upon us, Donoghue is one of those ushering it in. In 2001, he co-founded a company (Cyberkinetics) to develop and commercialize brain-computer interfaces. He and colleagues made headlines with a 2006 *Nature* report describing their work with Matthew Nagel, a young man paralyzed by a knife attack that severed his spinal cord. Surgeons implanted a 4 × 4–millimeter chip studded with 100 hair-thin electrodes into the part of Nagel's motor cortex responsible for planning arm movements. Now, as Nagel imagines moving his arm, a computer infers his intentions from the neural chatter. Videos accompanying the paper showed Nagel moving a cursor to operate an e-mail program and moving the fingers of a prosthetic arm. *Nature* apparently bleeped out Nagel's candid reaction when he first saw the hand respond to his thoughts: "Whoa, holy shit!" he says in an uncensored version Donoghue played at the meeting.

Three more patients, including one suffering from amyotrophic lateral sclerosis, have now received implants. All have been able to use the thought-controlled cursor without any training, Donoghue said. But video clips of their efforts showed that the cursor's movement is plodding and wobbly. When Nagel attempts to draw a circle onscreen, the result is subpar. "We're asking him to draw a circle with 24 neurons," Donoghue explains. "When we do something like that, we're using millions." Brown computer scientist Michael Black has developed algorithms to reduce the wobble—but so far the tradeoff is an even slower cursor.

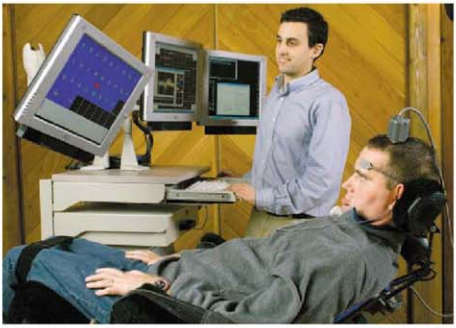
Other presenters described the potential for prosthetic devices for people deprived of hearing or sight. Stanford University Medical School otolaryngologist and surgeon Nikolas Blevins gave a brief history of research on cochlear implants, beginning with a seminal, if ill-advised, experiment by Italian physiologist Alessandro Volta circa 1790. Volta connected two metal rods to a battery and stuck them into his ear canals. Apparently unharmed, he reported hearing something like water boiling, thereby demonstrating that electrical stimulation could produce the sensation of sound.

Today's cochlear implants have restored hearing to tens of thousands of people but still have drawbacks. One of Blevins's patients, an articulate middle-aged woman with a trace of a British accent, said her implant "gave me back my life." But she still struggles to follow a conversation in a noisy restaurant and can't appreciate music. "It's just terrible, like honky-tonk piano or just bass and no melody," she said. The likely problem, Blevins said, is that individual nerve fibers in the cochlea normally respond to a narrow range of frequencies, but the electrodes in the implants stimulate many fibers at once.

Even so, cochlear implants are far ahead of retinal prosthetics, neuro-ophthalmologist Joseph Rizzo of Harvard Medical School in Boston told the audience. So far, about 50 people have received retinal implants, which transmit signals from a tiny camera to an array of electrodes attached to the retina. Patients tolerate the implants well, but exactly what they're able to see is difficult to know because the companies making the implants have been reluctant to release their data.

One of the most surprising exchanges occurred over dinner. Vilayanur Ramachandran of the University of California, San Diego, who had captivated the audience earlier with case studies of neurological curiosities, was about to tuck into his salad when he was interrupted by a tap on the shoulder. "Are you the guy who did that transgender study?" asked Stanford neurobiologist Ben Barres. He was. In a paper last year, Ramachandran hypothesized that transgendered people who have reassignment surgery might be immune to the "phantom penis" phenomenon. Just as many people who have had an arm amputated retain a vivid sense that the arm is still there, he explained, about 60% of men who have their penis amputated for cancer experience a phantom penis. He believes such sensations arise because the brain's representation of the body still has a place for the missing appendage. But does the brain's body representation include a penis for a woman born into a man's body? Ramachandran thought not, and a preliminary survey backed him up: Transgendered people were far less likely to report phantom penises (or breasts, in the case of female-to-male operations).

"That fits my experience exactly," said Barres, who is transgendered, adding that he'd heard similar stories from other transgendered people. Ramachandran seemed relieved. He said he'd gotten flak from some psychologist colleagues who didn't like his suggestion that some aspects of transsexuality could be explained by innate differences in the brain's body map. "I have a name for that," he said. "I call it neuron envy." **–GREG MILLER**



Think about it. Researchers are testing neural prosthetics that would enable paralyzed people like this man with ALS to control a cursor with their thoughts.

CREDIT: PHOTO BY MCKEEPHOTOGRAPHY.COM

BIOMEDICAL RESEARCH

The Graying of NIH Research

Many scientists who got their first grant in the 1950s or 1960s are still going strong. How do they view affirmative action for first-time grantees?

Roger Unger found himself drawn to research as a young internal medicine resident sometime around 1950, when he was treating diabetes patients in New York City. He had a controversial idea—that glucagon, a biomolecule then thought to be a contaminant in insulin made from ground-up beef and pork pancreases, might actually be a key hormone affecting blood sugar. Unger and colleagues in Texas had no direct evidence for this, but "we had the tools to answer the question, and we needed some money," Unger says. So at age 32, Unger applied for and won a research grant from the U.S. National Institutes of Health (NIH).

It didn't seem hard, "because I didn't know what I was doing back then," says Unger, now at the University of Texas (UT) Southwestern Medical Center in Dallas. Several years later, Unger's group published a landmark paper pinning down glucagon's role as a counter to insulin in regulating blood glucose levels: Glucagon tells cells to make more glucose, whereas insulin brings excess amounts down.

Today, at 84, Unger still runs a lab that enjoys NIH support. Now he's motivated by a new public health problem—the "meltdown" in Americans' health due to rising rates of obesity, he says. He's deep into exploring a concept his lab put forward: that a surfeit of lipids in obese people contributes to diabetes and heart disease. "I always decided I would retire when I ran out of ideas. But I didn't. The ideas got more exciting," says Unger.

That researchers such as Unger are still going strong in their 70s and 80s—and pulling down grants—would have been unheard of 3 decades ago. Because the biomedical enterprise was young and most universities had mandatory faculty retirement until 1994, there were few NIH-funded principal investigators older than 70 in 1980. But in 2007, there were at least 400 of them, according to NIH data. Indeed, NIH projections indicate that grantees over 68 could outnumber scientists under 38 by 2020 (see graph). The average age for obtaining a first NIH research grant is now 42. These data worry some research leaders, who have called on the community to reverse the trend. They have also contributed to a sense of crisis at NIH, which is taking steps to bolster the number of new investigators and slow the rising age of the average NIH-funded scientist, now 51 (see p. 834).

NIH officials say they do not mean to discourage very senior investigators from continuing in research. "It's not young against old," says NIH Director Elias Zerhouni. The number of investigators over 70 among those funded by NIH is a tiny fraction of the total, and some of them are incredibly productive into their later years—for example, Nobel laureates Eric Kandel and Paul Greengard are both around 80. Furthermore, peer review is supposed to winnow out any whose productivity has decreased. Scientists who have served on study sections generally say they haven't noticed a bias in favor of keeping older scientists' labs running, even if many of the reviewers are the applicants' former students and postdocs.

At the same time, concerns about the aging biomedical work force have prompted NIH to deploy what amounts to an affirmative action plan, setting numerical targets at each institute for grants to newcomers. To sample the community's views of this plan, particularly among those who won't benefit from the initiative, *Science* interviewed a score of researchers 70 or older. Most were drawn from a list of NIH investigators who have had the same basic NIH research grant, known as an R01, for at least 35 years, nearly all of them men. We asked: How does a very senior scientist decide when to shut down his or her lab? And does the current plight of young investigators influence their thinking? Most praised the idea of introducing fresh blood, but only about half said that they're ready to relinquish their own lab.

No time to quit

One strong theme—a sense that the review process was more interested in originality in the past—emerged in comments from this generation of scientists who applied for their first grants in the 1960s or earlier, often in their 20s or early 30s. It was a different game, they say. Not only did NIH have plenty of money to go around, but peer reviewers wanted ideas, not preliminary data. Microbiologist Samuel Kaplan, 74, of the University of Texas Medical School in Houston says he proposed studying a "newish" bacterium that he had never cultured. "If I submitted a proposal like that now, the study session couldn't stop laughing," he says. Peter von Hippel, 77, who earned a Ph.D. from the Massachusetts Institute of Technology in Cambridge at age 24 and then moved on to a postdoc there, found his grant waiting for him when he joined the Dartmouth Medical School faculty at age 28. "There was less to learn, and if you got on to a good project, things moved along pretty fast," says Von Hippel, now a professor emeritus and researcher at the University of Oregon, Eugene.

Some of these scientists were part of a

CREDITS: LEFT TO RIGHT: JASON KOSKI/CORNELL UNIVERSITY; KALMON ZABARSKY/BOSTON UNIVERSITY; UT SOUTHWESTERN MEDICAL CENTER

Lifelong passion. Harold Scheraga, 87, Phillips Robbins, 78, and Roger Unger, 84, are active researchers with NIH funding.

cadre who created the field of molecular biology. Others were pioneers in areas such as spectroscopy and protein chemistry. Forty or more years later, most have published hundreds of papers and trained scores of graduate students and postdocs. Many are members of the National Academy of Sciences. Some of those interviewed edit journals. (NIH intramural researcher Herbert Tabor, the editor of the *Journal of Biological Chemistry*, is nearly 90.) And many are still publishing in high-profile journals such as the *Proceedings of the National Academy of Sciences* and *Science*.

Most of those over 75 said they have cut back their research in recent years and stopped taking graduate students, who might be left in the lurch if their mentor developed health problems. Some have retired and are now emeriti, so their university no longer pays their salaries. Most say they are sympathetic to the funding difficulties faced by young investigators and support NIH's plans to target more grants to this group. "I couldn't agree more that we have to bring down the age of investigators," says Unger.

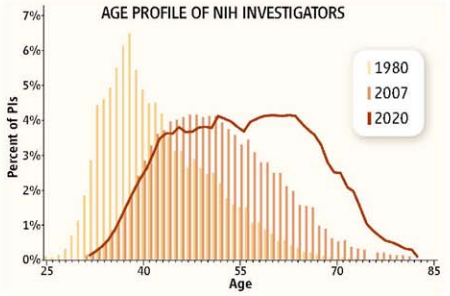
Representative of the nonretiring group is Cornell University protein chemist Harold Scheraga, 87, who may be the oldest NIH investigator. Since 1947, he has published more than 1200 papers, 20 of them in 2008. Scheraga is winding down an NIH grant for experimental work that expires in March 2009, which will free up lab space for a new faculty member, he points out. But he plans to continue with 10 workers on another NIH grant, the one that funds his theoretical study of protein folding, which he's had for 52 years.

"I'm very productive and making good progress," Scheraga says. "I'll keep going as long as I'm sane and my health is holding up. Only when somebody—my peers or myself—says that my science is washed up will I quit."

Some say that, like Unger, they're motivated by finding new research directions. The University of Pennsylvania's Robin Hochstrasser, 77, has a 14-person lab that is using lasers to study how protein structures change with time. "These techniques were only created 8 years ago. Close to 100 people are using them, and they started here," he says. But he praises NIH's new grants for young investigators and thinks setting targets for newcomer R01s is "reasonable … to ensure the future of medical research." John Dietschy of UT Southwestern, 76, says he has no plans to give up his R01 of 44 years on cholesterol metabolism, which ranked in the top 1% of proposals when it was last reviewed. "We're ahead of everybody in our field at the moment," says Dietschy. "As long as I'm having fun in the lab, we'll probably keep going."

The passion for doing research doesn't correlate with youth, some point out. "I think the people who are my age and continue to work in science have a certain amount of tenacity and they have a passion for it. I see the flame extinguishing in people in their 40s or 50s as much as people in their 70s," says molecular biologist E. Peter Geiduschek, 80, of the University of California (UC), San Diego. He says he will "keep doing research until somebody stops me from doing it." Like many others interviewed by *Science*, he says he can't imagine doing anything else.

Biochemist Carl Frieden, 79, of Washington University in St. Louis, Missouri, also says he will let peer reviewers tell him when it's time to close his lab. He says that although he's sympathetic to the struggles of young scientists, funding should be based strictly on scientific merit, not age. "We're the only profession judged by our peers every 3 to 5 years. If older scientists can pass that trial, I'm comfortable with that," he says.



Graying work force. NIH investigators are aging, and those over 68 could outnumber those under 38 by 2020.

SOURCE: NIH

Moving on

Others have decided to wind down. For molecular geneticist Robert Wells of Texas A&M University in College Station, who is just 70 but gave up his NIH grant and has been closing his lab for the past 2 years, the difficulties young investigators face were foremost in his thoughts. "If I and other old birds continue to land the grants, the [young scientists] are not going to get them," he says. He worries that the budget won't be able to support the 100 people "I've trained … to replace me." He will stay involved in science through advocacy.

Boston University biochemist Phillips Robbins, 78, has mixed feelings about his plans not to seek renewal of his grant when it ends in 3 years. He recently teamed up with a parasitologist to study glycosylation patterns in human parasites such as *Giardia*. "It's almost as though I've opened up a brand-new career," he says. But it's time, he says. "I think the folks who want to go out the door feet first, that that mindset is wrong. Once I reach 81, 82, it would be a poor decision for myself, for my university, and for students."

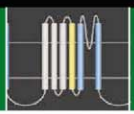
"It depends on what you're working on," suggests UC Berkeley biochemist Howard Schachman, 89, famous for fighting forced retirement at UC schools (*Science*, 14 September 1990, p. 1235). He says he let his last R01 lapse a few years ago only because his work studying a bacterial enzyme is out of date. "To what extent do you keep working and depriving young faculty of space in your department? I asked myself that question at 80 and decided I should keep going. But I couldn't do that today," he says. But Schachman, now emeritus, teaches the main biomedical research ethics course at Berkeley. "It's something that was interesting and important to me," he says.

For those researchers who do decide to leave the lab, the transition should be easier, says Harvard University molecular biologist Richard Losick, 65. He wishes there was more recognition for teaching and mentoring junior faculty members. "I don't think the culture of science fosters a graceful transition for aging scientists," says Losick, who says his own thoughts are to teach more in a few years. Others support the idea of giving retired faculty a small lab and encouraging them to keep up other activities.

Whatever their individual choices, the dilemma of how and when aging scientists should hang up their lab coats is only going to become more urgent. As Frieden points out, "It's rare to be as old as I, but there are going to be more of us."

–JOCELYN KAISER

With reporting by Rachel Zelkowitz.



LETTERS

edited by Jennifer Sills

Epigenomics: A Roadmap to Chromatin

IN THEIR LETTER "EPIGENOMICS: A ROADMAP, BUT TO WHERE?" (3 October, p. 43), H. D. Madhani *et al.* applaud the NIH for directing funds towards chromatin research, but argue that the Epigenomics Roadmap initiative (*1*) is ill-conceived and diverts funds from investigator-initiated proposals. However, their criticisms are more semantic than scientific, and they ignore the role that technology development has played in driving chromatin research. As recipients of grants awarded in this program, we would like to set the record straight.

We agree with Madhani *et al.* that epigenetic regulation is driven by transcription factor binding. However, studies of such regulatory processes have traditionally received strong NIH support, whereas the Epigenomics Roadmap aims to characterize the chromatin landscape that transcription factors act upon. Unlike transcription factors, which are diverse and often differ between eukaryotic taxa, chromatin components include histone variants and modifications that are essentially universal. Ultimately, transcription factors must act upon DNA packaged by histones, and essentially all eukaryotes use a common set of histone and DNA modifying enzymes, nucleosome remodelers, histone chaperones, and chromatin-binding proteins to facilitate transcription factor and polymerase action. We think that the NIH is justified in limiting this initiative to chromatin, and had some other term than "epigenomics" been used, there would be no basis for this complaint.

A more substantive concern is that this initiative diverts funds from investigator-initiated grants, corralling individual scientists "to work together under a more rigid, directed framework." However, 17 of the 22 grantees aim to develop novel tools and markers for chromatin research. Our three grants are high-risk, high-gain R21s; at $175,000 to $200,000 per year for 2 years, they are among the smallest NIH awards. Although we recognize that these funds might have been diverted from traditional programs of the NIH institutes that fund us (the National Institute on Drug Abuse, the National Institute of Diabetes and Digestive and Kidney Diseases, and the National Institute of Environmental Health Sciences), we believe that they deserve credit, not criticism, for investing in novel technologies for understanding chromatin.

STEVEN HENIKOFF,[1]* BRIAN D. STRAHL,[2] PETER E. WARBURTON[3]

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [2]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA. [3]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA.

*To whom correspondence should be addressed. E-mail: steveh@fhcrc.org

Reference

1. NIH Roadmap for Medical Research, Roadmap Initiatives, Epigenomics (http://nihroadmap.nih.gov/epigenomics/).

Bacteria by the Book

THE NEWS STORY "THE BACTERIA FIGHT back" (Special Section on Drug Resistance, G. Taubes, 18 July, p. 356) clearly describes the current resistance of bacteria to antibiotics and the diminishing pipeline of new products to treat them. However, Louis Rice mistakenly attributes the cause of the problem to physicians in past decades treating patients for "7 days, 10 days, 21 days, with no real reason other than making the doctor more comfortable," when the pneumonias "get better after 2 or 3 days."

In fact, physicians who treated pneumonias for longer periods were following established good medical practice. Medical textbooks from 30 years ago (*1*) recommended treatment of pneumonias for up to 3 to 4 weeks. Even today's medical textbooks (*2*) emphasize the need to treat patients with most bacterial pneumonias for 2 to 3 weeks. *Streptococcus pneumoniae* pneumonia should be treated for 1 to 2 weeks, *Mycoplasma pneumoniae* pneumonia for up to 3 weeks to avoid relapse, and Gram-negative bacilli pneumonia for "a minimum of 1 to 2 weeks" (*2*).

In addition, it was standard teaching in past decades that the administration of an antibiotic for fewer days than the full course would lead to the development of antibiotic resistance, since the surviving organisms of a partially treated infection would be selected for the presence of resistance.

EDWARD TABOR

Quintiles, 1801 Rockville Pike, Rockville, MD 20852, USA. E-mail: edward.tabor@quintiles.com

References

1. H. L. Barnett, A. H. Einhorn, Eds., *Pediatrics* (Appleton-Century-Crofts, New York, ed. 15, 1972), p. 1306.
2. L. Goldman, D. Ausiello, Eds., *Cecil Textbook of Medicine* (Saunders, Philadelphia, ed. 22, 2004), pp. 1768, 1773, 1776.

Response

TABOR CITES RECOMMENDATIONS IN RESPECTED textbooks of medicine and pediatrics regarding durations of antimicrobial therapy for common infections. There is no dispute regarding the existence of the recommendations. The issue is the evidence upon which these recommendations are based. In the News story, "The bacteria fight back" (Special Section on Drug Resistance, G. Taubes, 18 July, p. 356), I intended to point out that the available evidence actually supports the administration of short courses of antimicrobial agents for the treatment of pneumonia. Studies in the 1940s documented the excellent response of pneumococcal pneumonia to 2 to 3 days of therapy (*1*, *2*).

More recently, a European study compared 3 days with 8 days of amoxicillin therapy for community-acquired pneumonia and found that the results were excellent and equivalent for the two treatment durations (*3*).

Many current recommendations on therapy durations for common illnesses arose without experimental basis and during a time when antimicrobial agents were considered, at worst, a therapeutically neutral choice. We now recognize that there is no free lunch—administration of antimicrobial agents always comes at a cost of increased resistance and superinfections with serious pathogens such as *Clostridium difficile*. Given the ample data indicating that the risk of resistance increases with the duration of therapy (*4*, *5*), we as scientific and medical communities have an obligation to determine, in a scientifically credible way, minimal durations of therapy for common illnesses. The National Institutes of Health has acknowledged this reality by issuing a Broad Agency Announcement to support studies to look at optimal antimicrobial use (*6*).

The concept that shortened courses of therapy promote the emergence of resistance is a curious one. In vitro, susceptible organisms almost always out-compete resistant ones when there are no antibiotics around to provide a selective advantage (*7*). I suspect that this concept became established early in the antimicrobial era, as an explanation for relapses of resistant tuberculosis after short-course streptomycin therapy. The reality of the relapses was clear. The cause, however, was not the short duration of streptomycin therapy, but rather the existence of streptomycin-resistant subpopulations at the start of therapy that became the predominant population (*8*). We acknowledge this reality today by our use of combinations of antimicrobial agents, rather than longer courses of single agents, to treat tuberculosis. The concept that continued administration of an antibiotic will prevent emergence of a pathogen that is resistant to it simply makes no sense.

LOUIS B. RICE

Chief, Medical Service, Louis Stokes Cleveland VA Medical Center, 10701 East Boulevard, Cleveland, OH 44106, USA. E-mail: louis.rice@med.va.gov

References

1. M. L. Dawson, G. L. Hobby, *JAMA* **124**, 611 (1944).
2. W. S. Tillett *et al.*, *Bull. N.Y. Acad. Med.* **20**, 142 (1944).
3. R. el Moussaoui *et al.*, *Br. Med. J.* **332**, 1355 (2006).
4. J. Chastre *et al.*, *JAMA* **290**, 2588 (2003).
5. N. Singh, P. Rogers, C. W. Atwood, M. M. Wagener, V. L. Yu, *Am. J. Respir. Crit. Care Med.* **162**, 505 (2000).
6. National Institute of Allergy and Infectious Diseases, NIH, Targeted Clinical Trials to Reduce the Risk of Antimicrobial Resistance, Broad Agency Announcement (www.niaid.nih.gov/ncn/budget/concepts/c-mid0108.htm#03).
7. D. L. Andersson, *Curr. Opin. Microbiol.* **9**, 461 (2006).
8. F. Ryan, in *The Forgotten Plague: How the Battle Against Tuberculosis Was Won—and Lost* (Little, Brown, Boston, 1992), pp. 323–41.

Environmental Agencies: Lessons Learned

PREVIOUS ATTEMPTS TO ESTABLISH AN INDEPENDENT, science-based, environmentally focused federal research agency, including the National Biological Survey and the National Institute for the Environment, have ended in failure. Now we are treated to calls for "An Earth systems science agency" (M. Schaefer *et al.*, Policy Forum, 4 July, p. 44). Not surprisingly, the same objectives are being discussed: the need for a response-driven, flexible, interdisciplinary agency and the current poor organization, collaboration, and funding among existing agencies. This is undoubtedly true, but creating an umbrella superagency will not fix federal research programs.

In the U.S. Geological Survey (USGS), biological research has been treated as a poor stepchild to the earth sciences, and there is little doubt this will continue in the Earth Systems Science Agency (ESSA). Funding has been cut or shunted to administration; talented researchers have been drowned in paperwork, micromanaged, politicized, and subject to inept and incompetent leadership; and collaboration has become more difficult. It is ironic that some authors of the ESSA proposal oversaw procedures that have crippled biological research in USGS.

Changing the name and organization will not correct problems in federal natural science and environmental research. Unless the command and control mindset is lifted to free the creative energies of its scientists, federal science agencies will continue to lose their most valuable assets, scientists, as frustration and lack of support take a toll. Until scientists are recognized as valuable assets rather than full-time equivalents, the ESSA will be outmoded before the 10 years it takes to quell the administrative bickering.

C. KENNETH DODD JR.

Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32611, USA. E-mail: caretta@ufl.edu

Response

THERE MAY BE SOME TRUTH IN WHAT C. K. DODD writes, but history gives us important counter-examples. The creation of the National Oceanic and Atmospheric Administration (NOAA) in 1970 from existing agencies has led to better and stronger oceanic and atmospheric science, more applications, the sharing of a Nobel Prize, and numerous scientific awards. Today, there is a much stronger public and congressional awareness of environmental issues than at the time of the reorganizations he mentions.

Of course, not all of the bureaucracy, micromanagement, and command and control issues can be solved at once, and it is important to note that biology has been an equal partner in the problem faced by the entire U.S. Geological Survey (USGS)—a science organization in a budget-challenged management organization that has priorities other than science. We prefer to look on the positive side. The creation of an Earth Systems Science Agency (ESSA) provides the context to bring together synergistic subjects and researchers in an agency with a mission to provide societal benefits. We believe that this new context allows many, if not all, of the issues Dodd raises to be dealt with in an effective way so that researchers can prosper.

We have also emphasized the need for at least 25% of the agency's budget to be provided to grants, contracts, and cooperative agreements with academic and nonprofit institutions, in coordination with the National Science Foundation. It will be important that grant funding decisions are made in a forum where considerations from all parts of earth system science can be balanced. Outreach and links with outside institutions should also help the internal structure.

With respect to the biological sciences, ESSA would allow biodiversity and related issues to be examined in the context of both terrestrial and marine systems. Furthermore, ecosystem-based research and monitoring would be advanced by integrating atmospheric, oceanic, terrestrial, and freshwater biological and physical science programs.

Finally, although Dodd focuses on research, we note that our Policy Forum proposed a broader range for the new institution, reflecting the missions of both NOAA and USGS in research, monitoring, communication, and decision support systems that

inform policy-making and guide implementation. This broader mission benefits from the proposed reorganization and gives a framework in which research can be more easily facilitated (*1*).

D. JAMES BAKER,[1]* MARK SCHAEFER,[2] CHARLES F. KENNEL,[3] JOHN H. GIBBONS,[4] CHARLES G. GROAT,[5] DONALD KENNEDY,[6] DAVID REJESKI[7]

[1]djamesbaker@comcast.net; [2]markschaefer24@msn.com; [3]ckennel@ucsd.edu; [4]jackgibbons@hughes.net; [5]cgroat@mail.utexas.edu; [6]kennedyd@stanford.edu; [7]david.rejeski@wilsoncenter.org

*To whom correspondence should be addressed.

Note

1. Each of the coauthors has held senior Earth and environmental science positions in the federal government, including M. Schaefer: Deputy Assistant Secretary of the Interior, Acting Director of the U.S. Geological Survey; D. J. Baker: Administrator, National Oceanic and Atmospheric Association; J. H. Gibbons: Director, White House Office of Science and Technology Policy, Science Adviser to the President; C. G. Groat: Director, U.S. Geological Survey; D. Kennedy: Commissioner, Food and Drug Administration; C. F. Kennel: Associate Administrator, National Aeronautics and Space Administration, Director of Mission to Planet Earth; D. Rejeski: Served at the Office of Science and Technology Policy and Council on Environmental Quality from 1994 to 2000.

Homing In on a SIDS Model

IN THEIR REPORT "SPORADIC AUTONOMIC dysregulation and death associated with excessive serotonin autoinhibition" (4 July, p. 130), E. Audero *et al.* concluded that serotonin abnormalities play a role in sudden infant death syndrome (SIDS). However, the author omitted important discrepancies.

Although serotonin is ubiquitous in the autonomic system, and removing it would likely be catastrophic, studies have repeatedly shown that SIDS is not genetic, inherited, or even congenital (*1*). Furthermore, although there is adequate evidence of postmortem alterations in serotonin neurons in the brain stem, it is impossible to know whether these abnormal neurons were the cause or the result of SIDS (*2*). Audero *et al.* stated that mice expressing high levels of the serotonin receptor exhibited sporadic bradycardia and hypothermia. However, this claim is inconsistent with SIDS data. Bradycardia in SIDS is terminal, not just sporadic; bradycardia in SIDS has been shown to be associated with hypoxia (the study cited by Audero *et al.* examined only four infants and did not record oximetry) (*3*); and SIDS victims typically suffer from hyperthermia, not hypothermia (*4*). Finally, Audero *et al.* did not include published objections. WARREN G. GUNTHEROTH

Department of Pediatrics (Cardiology), University of Washington, Seattle, WA 98195, USA. E-mail: wgg@u.washington.edu

References

1. W. G. Guntheroth, P. S. Spiers, *Pediatrics* **110**, e64 (2002).
2. W. G. Guntheroth, P. S. Spiers, *JAMA* **297**, 1190 (2007).
3. R. G. Meny *et al.*, *Pediatrics* **93**, 44 (1994).
4. P. J. Fleming *et al.*, *Br. Med. J.* **313**, 191 (1996).

Response

WE AGREE WITH GUNTHEROTH THAT IT IS PREmature to call our *Htr1a*^{RO} line a mouse model of SIDS. However, we do think that our findings can contribute to research on SIDS. As Guntheroth points out, there is substantial evidence that changes in the medullary serotonin system are common in the brains of infants who died of SIDS, and may even be present in the majority of such infants. These findings have led to the hypothesis that changes in serotonin neurotransmission in the human medulla represent an underlying developmental defect that contributes to SIDS. However, despite the fact that the medullary serotonin system is well

known to play a role in autonomic physiology, until now no manipulations of the serotonin system, whether pharmacological or genetic, have been shown to precipitate death. In our opinion, the significance of our work lies in our discovering that a particular change in serotonin homeostasis, namely, increased autoinhibition, is linked to sporadic death. This finding suggests that further investigations into a possible role of altered serotonin autoinhibition in SIDS may be warranted.

Our data in no way endorse a purely genetic etiology for SIDS; potential changes in serotonin autoinhibition in at-risk infants are as likely to be genetic as environmental or stochastic. Neither do such changes have to derive from overexpression of Htr1a. As we mention in the Report, several other feedback circuits critical for serotonin homeostasis are known. Guntheroth is correct to point out that several features of our mice appear to be at odds with the clinical literature on SIDS. Our mice have excess Htr1a binding in the medulla, while SIDS brains have less; our mice show hypothermia and bradycardia before death, while existing data in infants is equivocal. In light of these discrepancies, we advise caution in treating the *Htr1a*^{RO} mice as a model of SIDS. Nevertheless, we believe that our mice offer a tool to understand how deficits in serotonin homeostasis can lead to death.

CORNELIUS GROSS* AND ENRICA AUDERO

Mouse Biology Unit, EMBL, Via Ramarini 32, 00015 Monterotondo, Italy.

*To whom correspondence should be addressed. E-mail: gross@embl.it

TECHNICAL COMMENT ABSTRACTS

Comment on "Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts"

Regis Debruyne, Carsten Schwarz, Hendrik Poinar

Gilbert *et al.* (Reports, 28 September 2007, p. 1927) reported that "hair shafts surpass comparably stored bone as an aDNA source [...] in regard to preservation and concentration of mtDNA." When experimental parameters are carefully controlled for, including adequate sampling, quantitative polymerase chain reaction analysis, and modeling the decay of DNA, the general importance of this claim is not supported.

Full text at www.sciencemag.org/cgi/content/full/322/5903/857a

Response to Comment on "Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts"

M. Thomas P. Gilbert, Webb Miller, Stephan C. Schuster

Debruyne *et al.* challenge the findings of our study and imply that we argue that hair shafts are an overall superior source of ancient DNA than bone. However, the authors are misreading and misinterpreting the conclusions of our study; we claim nothing further than that hair shaft represents an excellent source material for the shotgun sequencing of mitochondrial DNA genomes.

Full text at www.sciencemag.org/cgi/content/full/322/5903/857b

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

HISTORY OF SCIENCE

The Great American Pantheist

Jared Farmer

A Passion for Nature
The Life of John Muir
by Donald Worster
Oxford University Press, New York, 2008. 543 pp. $34.95, £18.99. ISBN 9780195166828.

In 2004, Governor Arnold Schwarzenegger selected the design of the California quarter in the U.S. Mint's current commemorative series. He chose an image of John Muir—identifiable by the biblical beard, walking stick, and rumpled suit—gazing at Yosemite Valley's Half Dome. According to the Austrian-born governor, the Scottish-born naturalist "has been a model for generations of Californians and conservationists around the world."

Yes, but a model of what? Muir's life was multilayered. Depending on which stratum one reads, Muir can be characterized as a model of amateur science, agrarian capitalism, or simple wanderlust. Today he is best known as the founding father of American environmentalism and most remembered for two periods of his life—the late 1860s and early 1870s, when he worked and wandered in the Sierra Nevada, and the 1890s and 1900s, when, as honorary president of the Sierra Club, he advocated for the protection and enlargement of Yosemite National Park.

CREDIT: UNITED STATES MINT IMAGE

In his comprehensive biography *A Passion for Nature*, Donald Worster shows Muir at every stage of life—a man in full, warts and all. We meet the draft-dodger who went to Canada during the Civil War and the domesticity-dodger who went to Alaska on season-long field trips during his daughters' childhood years. Like any good biographer, Worster (a professor of history at the University of Kansas) corrects the simplifications of popular memory. Readers may be surprised by certain details. For example, Muir was more gregarious than solitary. He made a small fortune managing an orchard staffed with Chinese laborers, about whom Muir felt race-based wariness. His writing career was facilitated by a series of wealthy benefactors, including Edward Harriman, a railroad magnate of the Gilded Age.

Worster also discusses Muir's contribution to geology. Muir lived in the proto-professional era of science. As a young man, he dreamed of following the footsteps of Alexander von Humboldt, combining personal discovery and scientific discovery while traveling to exotic tropical locales. In 1867, Muir embarked on a Humboldtian journey to South America. When a bout with malaria waylaid him in Cuba, he decided to go to California instead. There in the Sierra Nevada, during his off-hours as a sheepherder and sawmill operator, Muir joined the great scientific conversation of the day—breaking the biblical limits of time with geology and evolutionary theory. Without any institutional affiliation, Muir published in the proceedings of AAAS and corresponded with Louis Agassiz. Through his fieldwork, Muir made the case that the slow, uniformitarian work of glaciers—not some sudden, catastrophic event—created the sheer cliffs of Yosemite Valley.

Despite his ken for science, Muir lacked a scientific temperament. He was the opposite of disciplined and dispassionate. This came out when in 1877 Asa Gray and Joseph Hooker, two prominent champions of Darwin, climbed Mount Shasta with Muir. The august scientists wanted to talk pure science; they declined Muir's invitation to dance and shout, "Look at the glory! Look at the glory!" Gray and Hooker commented that "Muir is so eternally enthusiastic, we like to tease him." Whereas Gray famously tried to reconcile Darwinism with his belief in a Christian God—an earlier, more intelligent version of intelligent design—Muir advocated a more sacred yet less Christian position. After abandoning the Calvinism of his father, Muir developed a concept of "God" synonymous with beauty and harmony—universal principles of nature. In a clever turn, Worster employs Linnaean taxonomy to describe Muir's belief system: *Pantheism muirii* var. *sierra*. In today's world when science so often gets dragged into bipolar debates between theists and atheists, Muir offers a historical example of a third way. He felt equally comfortable with the language of science and the language of religion. For him, holism was a spiritual as well as a scientific pursuit.

It is ironic, then, that Muir's ecological sensibility—his holistic view of biological systems—was stunted. As Worster points out, Muir's fixation on mountain geology and mountain scenery blinded him to the ecological importance of unspectacular lowland environments like wetlands. Muir privileged faraway wilderness areas over local inhabited spaces. After Yosemite, his favorite landscapes were the glacial bays and fjords of Alaska. Similarly, he focused his botanical enthusiasm on individual sublime species like the giant sequoia (*Sequoiadendron giganteum*) and the coast redwood (*Sequoia sempervirens*). Late in life, when he traveled across the globe, he went looking for other champion trees such as Australia's mountain ash (*Eucalyptus regnans*) and Africa's baobab (*Adansonia digitata*). It has taken the environmental movement a long time to overcome the Muirian bias for extraordinary nature. Without faulting Muir personally, it seems fair to say that biodiversity would have been better served had the Sierra Club been complemented by a Marsh Club, a Prairie Club, and a Desert Club—not to mention an Urban Nature League.

Worster clearly admires his subject and even speculates that Muir's life may demonstrate E. O. Wilson's biophilia hypothesis. Perhaps. (If we need prophets like Muir to remind us of our innate passion for nature, how innate can it be?) It is remarkable that Worster, an environmental historian who has been typed as a "declensionist"—someone who focuses unremittingly on how humans have degraded nature—displays such Muir-like faith in the transformative power of nature worship. He insists that we still have much to learn from the great American pantheist. A radical egalitarian, Muir argued for the natural rights of other living things. A radical optimist, Muir believed that industrial capitalism and nature preservation could be reconciled.

In one of his most quoted passages, Muir condemned those who would—and did—build a dam inside Yosemite National Park: "These temple destroyers, devotees of ravaging commercialism, seem to have a perfect contempt for Nature, and, instead of lifting their eyes to the God of the mountains, lift them to the Almighty Dollar." Compared to most of Muir's writing, that passage is unusually political and priggish. Muir's full life demands a different

The reviewer is at the Department of History, State University of New York, Stony Brook, NY 11794–4348, USA. E-mail: jared.farmer@stonybrook.edu

kind of epitaph—something more impure and for that reason more uplifting. One suspects that Muir might actually agree with Schwarzenegger, who, standing beside the quarter-dollar image of the apostle of nature, said, "Here in California, growth, progress, wilderness protection and the protection of the environment go hand in hand, even though some people believe that you can only have one or the other, we want to be committed to make it go hand in hand."

10.1126/science.1166796

HISTORY OF SCIENCE

The Kingdom of Plants

John C. Waller

When Joseph Dalton Hooker returned to England in 1843 from a Royal Navy expedition sent to the South Magnetic Pole, he could only lament the state of British botany. Leading scientists were talking of the decline of science as a result of state parsimony, but the outlook for aspiring botanists like Hooker was particularly bleak. The discipline to which he wished to devote his life remained a poor relation of the physical sciences or zoology. Its practitioners had made few of the bold and brilliant generalizations that marked out a science as being truly "philosophical," while gardeners and amateur collectors sporting trowels, bags, and prettily illustrated handbooks laid as much claim to the title "botanist" as did the head of a vast herbarium like Kew Gardens in west London. In fact, even Kew was still making the transition from royal park to a state-funded center of botanical research.

Hooker deplored the lowly status of botany, and he had strong personal reasons to want to hurry it into a state of maturity. Intelligent, educated, and well-traveled, Hooker was also out of pocket. Even when he had gained a salaried position at Kew, the rewards for his labors were only modest. And, to make matters worse, he realized that many of his contemporaries considered the pursuit of pure knowledge to be sullied by earning a wage.

The reviewer is at the Department of History, 301 Morrill Hall, Michigan State University, East Lansing, MI 48824, USA. E-mail: wallerj1@msu.edu

In *Imperial Nature*, Jim Endersby shows how the person who was to become Britain's foremost botanist and Darwin's right-hand man strove to get botany invited to the high table of Victorian science. In doing so, Endersby focuses on the practical dimensions of Hooker's drive to establish the reputation of Victorian botany: how he obtained properly preserved specimens from far-flung regions of the globe, reconciled his gentlemanly status with drawing a wage from doing science, and sought to make botany economically useful to his nation by having trees and plants (such as rubber and sisal) shipped from one part of the empire to another.

Endersby's story is as much about the exercise of power as the acquisition of legitimate expertise: scientific advance and self-interest went hand in hand as Hooker and his allies elevated botany to a higher plane. Accordingly, several chapters follow Hooker in his dogged attempts to assert the primacy of metropolitan botanists like himself over a multitude of amateur enthusiasts and colonial collectors. To this end, isolated colonial collectors, many of them incorrigible taxonomic "splitters," were told that they lacked the broader perspective needed to say where one species ended and another began; only metropolitan experts had access to the extensive herbaria and libraries necessary for conducting proper systematics. Similarly Hooker and company claimed sole authority to name the empire's plants, to the disappointment of collectors like William Colenso in New Zealand, who would have preferred to use Latinized versions of Maori terms for the specimens he sent to Hooker at Kew. Collectors had to be kept sweet, but they were still taught to see themselves as "worker bees." Hooker keenly resisted the attempts of some of his collectors to indulge in theorizing, arguing that those armed with only local knowledge were unfitted to grapple with the bigger, abstract issues. And it was theorizing, for Hooker, that would make botany into a recognizably scientific endeavor, allowing the metropolitan expert to move beyond dry lists of species and genera.

Endersby (a historian of science at the University of Sussex) astutely reveals the difficulties of the relationship between metropolitan botanist and colonial collector. And his book usefully reminds us that underpinning many of the advances in theory made by naturalists of the 1800s were the efforts of vast networks of these collectors. Without the often-unpaid work of those who labored in jungles, forests, and marshes; on mountains; and along shores in search of rare examples of fauna and flora, neither Hooker nor Charles Darwin could have made the breakthroughs they did.

Imperial Nature
Joseph Hooker and the Practices of Victorian Science
by Jim Endersby
University of Chicago Press, Chicago, 2008. 441 pp. $35, £18. ISBN 9780226207919.

Fashion setter. Hooker's *Rhododendrons of Sikkim Himalaya* (1849) led many British gardeners to grow the genus for its colorful blossoms.

Endersby also argues that the directors of herbaria, like Kew, were inclined to be taxonomic "lumpers" rather than splitters in part because they would otherwise have been overwhelmed by the sheer number of plants to classify. This brings us to Hooker's relationship with Darwin. One of the reasons that Hooker became a Darwinian, says Endersby, is that the theory of natural selection chimed with his own preference for lumping. Darwin's emphasis on variability allowed Hooker to insist that naturalists must not let slight differences between one plant and another mislead them into erecting more and more species categories.

Imperial Nature is not a conventional scientific biography. The usual fare of birth, love, and death is largely absent. Instead, Endersby give us a detailed, scholarly account with a deeper point: that science is about more than the grand battles of competing ideas. In doing so, he provides a richly textured account of a period in which the status of natural science was far more precarious than it is today. And the book will hopefully stand as a reminder, during next year's Darwin celebrations, of just how many unsung individuals contributed to the scientific progress of the age.

10.1126/science.1165855

CREDIT: PLATE 30 FROM *RHODODENDRONS OF SIKKIM HIMALAYA* (1849)/NATURAL HISTORY MUSEUM, LONDON/ALAMY

GENETICS

The Human Variome Project

Richard G. H. Cotton,[1,2,3]* Arleen D. Auerbach,[1] Myles Axton,[1] Carol Isaacson Barash,[1] Samuel F. Berkovic,[4] Anthony J. Brookes,[1] John Burn,[1] Garry Cutting,[1] Johan T. den Dunnen,[1] Paul Flicek,[1] Nelson Freimer,[5] Marc S. Greenblatt,[1] Heather J. Howard,[2] Michael Katz,[1] Finlay A. Macrae,[1] Donna Maglott,[1] Gabriela Möslein,[1] Sue Povey,[1] Rajkumar S. Ramesar,[1] Carolyn S. Richards,[1] Daniela Seminara,[1] Timothy D. Smith,[2] María-Jesús Sobrido,[6] Jan Helge Solbakk,[1] Rudolph E. Tanzi,[7] Sean V. Tavtigian,[1] Graham R. Taylor,[1] Joji Utsunomiya,[1] Michael Watson[3]

An ambitious plan to collect, curate, and make accessible information on genetic variations affecting human health is beginning to be realized.

It has been 60 years since the first variation causing inherited disease was defined at the protein level. Currently, at least one such mutation is known to have occurred in 3000 of the 20,000 recognized human genes. In the next few years, the number of genes in which disease-causing mutations are recognized will increase dramatically. Despite good intentions, efforts to develop and build databases have failed to keep up with this pace.

Thus, clinicians and diagnostic laboratories must waste their time trawling through many publications and databases to determine whether a mutation found in a patient has been previously characterized. Availability of previous characterizations of all mutations and their effects would allow them to base their diagnoses and prognoses on evidence rather than guesswork and conjecture. For inherited diseases, rapid access to curated information on all mutations in all genes from all populations is needed. Note that those who gain most by the availability of up-to-date gene variant data are usually downloading information only and are failing to add their findings to further improve the quality of the data collected. Changing this attitude and collecting all data seem to be mammoth tasks, but they are essential.

The Human Variome Project (www.human-variomeproject.org/), initiated in 2006 (*1*–*3*) is the global community's effort to collect, curate, and make accessible information on all genetic variations affecting human health. The specific objectives are to encourage the development and adoption of standards; define, reach consensus on, and implement ethical requirements (including informed consent forms and approaches for protecting patient confidentiality); develop systems for automated data submission; develop community education and communication programs; enable participation by developing countries; support curation processes; promote evidence-based genetic medicine; and create usable systems for contribution, curation, search, and retrieval.

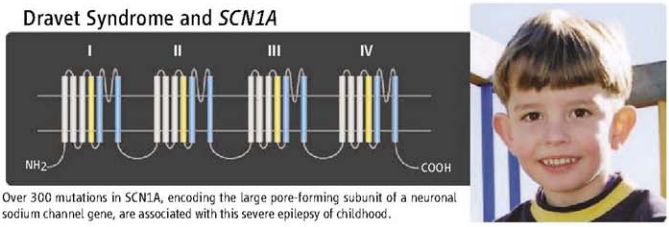
The Human Variome Project is complementary to the massive resequencing projects that contribute on a daily basis to variation databases such as dbSNP and the increasing information from genomewide association studies. These large data sets provide excellent population data on all variations compatible with life and also variations associated with common diseases. But ascertainment of mutations through observation of rare phenotypes provides a quality of information that will not otherwise be captured.

Proofs of Principle

Recent progress indicates that the collection of mutations in all genes worldwide is possible. For example, the International Society for Gastrointestinal Hereditary Tumours (InSiGHT; www.insight-group.org), since February 2007, has embarked on developing a pipeline to collect legacy and new data, to place all data from disparate databases on the Leiden Open (source) Variation Database (*4*) (LOVD; www.lovd.nl) with freely available software and to develop submissions to National Center for Biotechnology Information (NCBI, U.S. National Institutes of Health); European Bioinformatics Institute (EBI); and University of California at Santa Cruz (UCSC) Genome Browser. InSiGHT has begun a pilot project for the collection of all mutations from all countries in four mismatch repair genes that are altered in colon cancer patients. Strategies need to be developed and appropriate software created and put in place in the next 3 years that will enable the seamless, effortless, and low-cost collection of data from laboratories, clinics, and hospital records and their delivery to appropriate databases.

One project that has just begun represents a creative approach to funding for supporting data curation. This cost varies depending on the extent of work involved and the number of mutations per gene, but estimates range from $1000 to $200,000 per year. The Adopt-a-Gene Program through the Human Variome Project is encouraging industry and patient support groups to sponsor the curation of specific genes. The first of these partnerships is already under way, with CMO Global Services supporting the Familial Hemiplegic Migraine Variation Database (http://lovd.nl/FHM).

When a variation is found in a patient, the



Over 300 mutations in SCN1A, encoding the large pore-forming subunit of a neuronal sodium channel gene, are associated with this severe epilepsy of childhood.

PHOTO CREDIT: INGRID SCHEFFER

[1]Discussion leaders for the Human Variome Project Planning Meeting 2008. [2]Genomic Disorders Research Centre, Howard Florey Institute, Melbourne, Australia. [3]Cochair of the HVP Planning Meeting. [4]Epilepsy Research Centre. [5]University of Melbourne, Austin Health West Heidelberg, Australia. [6]UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA. [7]Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain, and Center for Network Biomedical Research on Rare Diseases (CIBERER), Institute of Health Carlos III, Madrid, Spain. [7]Harvard Medical School & Genetics and Aging Research Unit, MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Charlestown, MA, USA. Complete affiliations are listed in the supporting online material.

*Author for correspondence. E-mail:cotton@unimelb.edu.au

clinician or diagnostician has to decide whether it is causing the disease. Numerous algorithms have been developed that predict pathogenicity on the basis of such features as evolutionary conservation, frequency, nature of missense change, and protein structural changes. Ideally, an algorithm should incorporate all of these features to give a probability of pathogenicity. International efforts to create such an algorithm are ongoing. Below, we describe efforts to create the next generation of databases for one group of disorders.

Neurological Disorders

The brain has the largest number of expressed transcripts of all organ systems, and this is reflected in the very large number of neurogenetic disorders. Over 30% of Mendelian diseases have neurological manifestations (*5*). When the molecular lesion is singular and easily detected, such as the triplet repeat expansion in Huntington disease, molecular testing has revolutionized clinical practice. Accurate diagnosis and genetic counseling can now be given.

However, the situation is more complicated in most neurogenetic diseases. Neurogenetics is full of examples of genetic heterogeneity (e.g., Charcot-Marie-Tooth disease); allelic disorders (e.g., hemiplegic migraine and episodic ataxia); variable expressivity (e.g., synucleinopathies); incomplete penetrance (e.g., dopa-responsive dystonia); anticipation (e.g., spastic paraparesis); phenocopies (e.g., familial epilepsies); and imprinting (e.g., Angelman and Prader-Willi syndromes), all of which are biological phenomena that entangle genotype-phenotype relationships. To further complicate matters, mitochondrial mutations add their particular inheritance mechanism and heterogeneous phenotypic expression [e.g., Leigh syndrome and Neuropathy, Ataxia, and Retinitis Pigmentosa (NARP)].

Neurological phenotypes are extremely varied and often complex and can evolve over time in a given patient. Two individuals with the same genetic and pathologic process may show different phenotypes, whereas the same phenotype (e.g., lack of fine motor coordination) may be manifested in totally different pathogenic events and disorders.

A particularly good example in which a complete catalog of variation would be invaluable is in a severe epilepsy, Dravet syndrome (see figure on page 861). It begins at 6 months of life with febrile seizures, followed by later intellectual decline and a somber prognosis. Some evidence suggests that early aggressive treatment may improve outcome, making early diagnosis imperative. The disease is associated in ~80% of cases with variation in the neuronal sodium channel subunit gene *SCN1A;* a large gene with 26 exons. Half the cases have mutations predicting truncation of the protein, but half are missense. However, there is a milder epilepsy syndrome [generalized epilepsy with febrile seizures + (GEFS+)] that begins at the same time (with febrile seizures) for which treatment is often not required, the outcome is good, and missense mutations in *SCN1A* are found in about 10% of cases. Our knowledge base is insufficient to predict the phenotype associated with particular missense mutations in *SCN1A*.

A number of public neurological databases already exist, such as the Alzheimer Disease and Frontotemporal Dementia Mutation Database (www.molgen.ua.ac.be/ADMutations) and the Inherited Neuropathies Database (www.molgen.ua.ac.be/CMTMutations). However, they generally contain limited data on the phenotype and on the evidence of pathogenicity. Phenotype-based databases, on the other hand, such as the London Neurogenetics Database (*6*) or GeneReviews (www.geneclinics.org/) are detailed in disease description but usually not in the interpretation of specific mutations and variations in a particular gene.

Among the challenges to meet for neurogenetic databases is that of finding appropriate forums and funding policies that allow the confluence of "genotypers" and "phenotypers," i.e., multidisciplinary teams that include molecular geneticists, cell physiologists, and biochemists, as well as clinicians (adult and pediatric neurologists, clinical neurophysiologists, and neurosurgeons).

What organizations are best suited to call for convocation of such multidisciplinary teams? International networks for neurological disorders such as spinocerebellar ataxias, Charcot-Marie-Tooth disease, or spastic paraparesis could facilitate collection of genetic and clinical information from basic science and clinical research groups. However, the information generated by these consortia often remains available only to members, with important legal and ethical questions emerging from the use of these large patient data sets and sample collections. Additional limitations of disease-centered consortia are their dependence on financial support for a specific timeframe, marked by funding or mission views.

Alternative forums might be neurogenetics societies; however, the current landscape of neurogenetics meetings still needs to evolve to face the multidisciplinary effort involved. There are neurogenetics study groups within some neurological societies, such as the Spanish Society for Neurology (SEN), the American Academy of Neurology (AAN), or the European Federation of Neurological Societies (EFNS). However, these panels are generally small, run by neurologists with an interest in genetics, and with very limited participation of geneticists and basic researchers. Associations for human genetics generally lack specific neurogenetics committees, plus clinical neurologists are absent or minimally represented in these genetic forums. (One exception is the German Society for Neurogenetics.)

An example of the difficulties can be found in Spain, where nine large networking national centers for research have recently been created by the Spanish Health Institute Carlos III (CIBERs), and are responsible for coordinating translational biomedical research in specific areas. Among the CIBERs, the Center for Research on Rare Diseases (CIBERER) has established a neurogenetics committee, but virtually no clinical neurologists, neurophysiologists, or neurosurgeons belong to the CIBERER. Another CIBER on neurodegenerative disorders (CIBERNED) includes basic neuroscientists, pathologists, and clinical neurologists, but only one or two clinical diagnostic geneticists. Any initiative aimed at organizing the data on genetic neurological disorders in Spain would have to bring these groups together. There is currently an initiative to launch a Neurogenetics Association in Spain.

A strategy for the development of neurological locus-specific databases could start with international, multidisciplinary, disease-centered networks. These would be organized into a multidisciplinary neurogenetics society or network with representatives of each disease consortium and/or locus-specific database. They would integrate the information, propose common guidelines, discuss common coding issues, and facilitate navigation from one database to another. The role of the Human Variome Project will be to foster such a strategy, to get disease-centered networks involved, and to promote or host coordination forums.

References and Notes

1. Editorial, *Nat. Genet.* **39**, 423 (2007).
2. H. Z. Ring, P. Y. Kwok, R. G. Cotton, *Pharmacogenomics* **7**, 969 (2006).
3. R. G. Cotton *et al.*, *Nat. Genet.* **39**, 433 (2007).
4. I. F. Fokkema, J. T. den Dunnen, P. E. Taschner, *Hum. Mutat.* **26**, 63 (2005).
5. T. Costa, C. R. Scriver, B. Childs, *Am. J. Med. Genet.* **21**, 231 (1985).
6. M. Baraitser, R. M. Winter, *London Dysmorphology Database, London Neurogenetics Database & Dysmorphology Photo Library on CD-ROM* (Oxford Univ. Press, Oxford, 2001).

7. We thank I. Scheffer and the patient's family for the photograph.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/861/DC1

10.1126/science.1167363

BIOCHEMISTRY

Getting Close to Termination

The action of release factors on ribosomes has now been clarified by crystallography.

Anders Liljas

Ribosomes are complexes of RNA and proteins that constitute the sites of protein synthesis in cells. Messenger RNA (mRNA), transcribed from genomic DNA, is translated into protein through a decoding process that involves ribosomes, transfer RNAs (tRNAs) that carry amino acids, and other parts of the cellular machinery. Amino acids are thus assembled into a polypeptide chain according to a sequence that is encoded by the corresponding gene. Two papers—one by Weixlbaumer *et al.* on page 953 of this issue (*1*), and the other by Laurberg *et al.* in *Nature* (*2*)—now report crystal structures that shed light on how translation is terminated and how the finished protein is released from the ribosome. The studies investigate two closely related proteins called release factors that respond to different stop codons (nucleotide sequences in mRNA that signal termination of translation) in bacteria.

Translation of mRNA into protein by ribosomes proceeds through initiation and elongation of the polypeptide chain until a stop codon in the mRNA arrives at the decoding site (or A site) of the ribosome. Usually, there is no tRNA that binds to the stop codon and ribosome at this point. However, either of two release factors can hydrolyze the nascent polypeptide from the tRNA that it is bound to the P site of the ribosome (adjacent to the A site) with high accuracy (*3*). These two release factors, called RF1 and RF2, both respond to the stop codon encoded by the nucleotides UAA. RF1 also uniquely interacts with UAG and RF2 with UGA. The release factors both identify stop codons on the small subunit of the ribosome, as well as hydrolyze the peptide off from the tRNA in the peptidyl transfer center on the large subunit of the ribosome. Thus, the release factors span the same distance, about 75 Å, as the tRNAs do on the ribosome (see the figure).

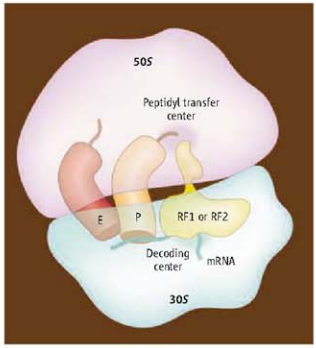
Studies of the release factors have identified a region in these proteins that switches their stop codon specificities. Thus, exchange of the tripeptide Pro-Xxx-Thr (where Xxx is any amino acid) of RF1 for the tripeptide Ser-Pro-Phe of RF2 flips their codon specificities (*4*). The structure of the release factors alone showed a closed form of the proteins, with a distance of only 23 Å from the Gly-Gly-Gln sequence (essential for the hydrolysis reaction that frees the polypeptide from the tRNA) to the part of the factor that is important for codon specificity (*5*). Cryo-electron microscopy and crystallographic investigations showed that the release factors undergo large conformational changes upon binding to the stop codons on the ribosome (*6*–*8*).

The current papers provide the crystal structures of *Thermus thermophilus* RF1 (*1*) and RF2 (*2*) bound to ribosomes (*9*) from the same species, at 3.2 and 3.45 Å resolution, respectively. RF1 interacts with the stop codon UAA and RF2 with UGA. Part of each release factor occupies essentially the position of a tRNA in the A site of the ribosome. The key questions are how the two factors decode the stop codons and how the Gly-Gly-Gln sequence participates in introducing a water molecule into the peptidyl transfer center to release the polypeptide chain from the tRNA through hydrolysis.

The universally conserved nucleotides (G530, A1492, and A1493; *Escherichia coli* numbering) in the 16*S* RNA, which is part of the small ribosome subunit, undergo codon-anticodon interactions. This stabilizing conformation would prevent release factors from binding. Instead, a different arrangement of these nucleotides is observed.

The stop codons begin with a U (U1), which interacts with two conserved glycines on the release factor. Nucleotides A and G, (which are larger than U and C) in this first position of the stop codon would clash with a release factor. A C nucleotide (the same size as U) could not make hydrogen bonds with a release factor. For RF1, the threonine of its Pro-Xxx-Thr motif does hydrogen bond with U1. Furthermore, both the proline and threonine interact with A in the second position of the stop codon (A2). The threonine side chain in the release factor motif that donates its hydrogen to U1 can only function as a hydrogen bond acceptor for A2. This may explain why RF1 does not bind to a G in this position. In RF2, the second nucleotide base can be A or G. The serine of the Ser-Pro-Phe motif can hydrogen bond to either G2 or A2. RF1 and RF2 could not form direct hydrogen bonds with the nucleotides U and C. The third base of the stop codon does not "stack" on the nucleotides in position 1 and 2, but on G530 of the 16*S* RNA. Nucleotides A and G are selected by the release factors. The stacking of A or G on G530 is more favorable than for C and U. In addition, hydrogen bonds participate in the selection of A or G. Thus, the tripeptide motifs of the release factors are important for decoding stop codons, but are not solely responsible for the specificity of the interactions.

The two crystal structures describe the state after the peptide has been released from the ribosome. Nevertheless, attempts to describe the catalytic mechanism can be made. The conserved Gly-Gly-Gln motif in the release factor involved in peptide release is located in a loop of the protein and when in complex with the ribosome, it is positioned in the peptidyl



Release from the ribosome. The *T. thermophilus* 70*S* ribosome is shown with mRNA (green) and two tRNA molecules bound to the E site (red) and P site (orange) of the ribosome, respectively. Due to the presence of a stop codon in the decoding center, a release factor of class 1 (RF1 or RF2) binds to the A site. The release factor has an open conformation. Interaction between the stop codon in the mRNA and the release factor tripeptide at the decoding center of the ribosome is now clarified (*1*, *2*). At the opposite end of the release factor, the universally conserved Gly-Gly-Gln motif participates in hydrolyzing the peptide from the peptidyl-tRNA in the P site. This process occurs in the peptidyl transfer center of the ribosomal large subunit.

Department of Molecular Biophysics, Lund University, Sweden E-mail: anders.liljas@mbfys.lu.se

transfer center. The two glycines adopt conformations that are not possible for other amino acids, which may explain why they are conserved; if they are mutated to other residues, the activity is dramatically decreased (*10*, *11*). If other residues replace the glutamine, the factor does not function in vivo, but in vitro the effects are less dramatic. Also, the amino group of the glutamine is methylated, which increases the rate of termination.

The Gly-Gly-Gln loop would clash with ribosomal RNA nucleotides that protect the ester bond between the tRNA and the nascent peptide. When ribosomal RNA nucleotides are pushed away by the release factor, the ester bond is exposed to nucleophilic attack by a water molecule. The main-chain amide of the glutamine makes a hydrogen bond with the tRNA that sits at the P site of the ribosome. The glutamine side chain may coordinate the water molecule involved in peptide hydrolysis. A molecular dynamics study also suggests such a location (*12*). Furthermore, the functional importance of the methylation may suggest that the glutamine interacts directly with the water molecule (*13*).

Characterization of the roles of the two tripeptides in release factors that decode stop codons in mRNA brings us closer to understanding termination of protein synthesis. As well, we now have a converging picture of the role of the universal Gly-Gly-Glyn peptide in hydrolysis of the ester bond between tRNA and the growing polypeptide in protein synthesis.

References

1. A. Weixlbaumer *et al.*, *Science* **322**, 953 (2008).
2. M. Laurberg *et al.*, *Nature* **454**, 852 (2008).
3. D. V. Freistroffer, M. Kwiatkofsky, R. H. Buckingham, M. Ehrenberg, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2046 (2000).
4. K. Ito, M. Uno, Y. Nakamura, *Nature* **403**, 680 (2000).
5. B. Vestergaard *et al.*, *Mol. Cell* **8**, 1375 (2001).
6. U. B. Rawat *et al.*, *Nature* **421**, 87 (2003).
7. B. P. Klaholz *et al.*, *Nature* **421**, 90 (2003).
8. S. Petry *et al.*, *Cell* **123**, 1255 (2005).
9. M. Selmer *et al.*, *Science* **313**, 1935 (2006).
10. A. V. Zavialov, L. Mora, R.H. Buckingham, M. Ehrenberg, *Mol. Cell* **10**, 789 (2002).
11. J. J. Shaw, R. Green, *Mol. Cell* **28**, 458 (2007).
12. S. Trobro, J. Åqvist, *Mol. Cell* **27**, 758 (2007).
13. V. Heurgue-Hamard, S. Champ, Å. Engström, M. Ehrenberg, R. H. Buckingham, *EMBO J.* **21**, 769 (2002).

10.1126/science.1166913

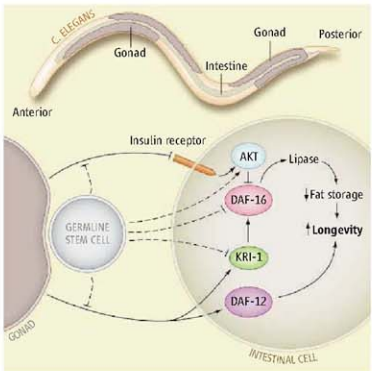
PHYSIOLOGY

Burn Fat, Live Longer

Endocrine signals from germline stem cells control fat metabolism in the worm, thus affecting the animal's life span.

Ting Xie

Genetic factors and reproductive and nutritional status influence aging. Most of the progress toward understanding the underlying molecular mechanisms comes from genetic studies on model organisms such as the worm *Caenorhabditis elegans* and the fly *Drosophila melanogaster*. These studies have revealed important roles for reproduction, dietary (caloric) restriction, stress, and the hormone insulin in regulating aging (*1*–*4*). Recently, germline stem cells—those in the gonad that can renew and also differentiate into gametes—in the worm and fly have been shown to produce a signal that enhances aging, whereas somatic cells of the gonad produce a signal that retards aging (*5*, *6*). This coupling of reproduction and aging may have maximized the use of natural resources for reproduction during evolution. Although dietary restriction is an important means for extending life span in many species (*2*), it remains unclear whether aging is directly attributable to altered fat metabolism that is induced by a restricted diet. The study by Wang *et al.* (*7*) on page 957 of this issue may provide some insights into this question, as it reveals fat metabolism as an unexpected link between reproduction and aging.



A model for longevity. In *C. elegans*, somatic cells in the gonad send signals (currently unknown) to intestinal cells that reduce fat storage through pathways involving KRI-1–DAF-16 and DAF-12. The result is increased longevity. Germline stem cells inhibit these pathways. Signaling through the insulin receptor also controls longevity by regulating expression of a triglyceride lipase, and thus fat storage, through a DAF-16–dependent pathway. This pathway might also be involved in the regulation of longevity in response to germline stem cell depletion.

In organisms ranging from worm to human, the insulin/insulin-like growth factor signaling (IIS) pathway is highly conserved and controls longevity (*8*). In *C. elegans*, a defective IIS pathway due to mutations in *daf-2* (which encodes an insulin receptor) and in other pathway components leads to nuclear localization of the transcription factor FOXO/DAF-16 and to longevity. By contrast, constitutive signaling by this pathway prevents nuclear localization and shortens life span (*9*, *10*). Germline stem cells in *C. elegans* directly control aging: Animals depleted of these cells are long-lived, whereas their overproliferation results in short-lived worms (*5*). Genetic screens of the worm have identified *kri-1* (which encodes an ankyrin-repeat protein), *daf-12* (which encodes a nuclear hormone receptor), and *daf-16* as required for extending life span in response to germline stem cell depletion, suggesting that a signal (likely hormonal) released from cells of the reproductive system controls longevity (*11*). A defective IIS pathway causes the nuclear localization of FOXO/DAF-16 in multiple cell types, whereas the depletion of germline stem cells causes nuclear location primarily in intestinal cells (*11*). Thus, intestinal cells are the likely targets of the reproductive signals. Wang *et al.* further identified molecular targets of the reproductive signals in intestinal cells (*7*).

Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA. E-mail: tgx@stowers-institute.org

Previous studies have implicated adipose tissue and the reproductive system in the regulation of longevity in the fly and the mouse (*3*, *4*, *12*–*14*). Contrary to the prediction that germline stem cell proliferation is expected to consume most of the energy and to reduce fat storage in *C. elegans*, Wang *et al.* observed that germline stem cell depletion—either by laser ablation of germline stem cell precursors or by a mutation that blocks a pathway that supports stem cell self-renewal (involving the signaling molecule Notch-like receptor GLP-1)—caused the reduction of lipid droplets in intestinal cells where fat normally is stored. On the other hand, overproliferation of germline stem cells—by blocking their differentiation through constitutive GLP-1 signaling—increased fat storage in intestinal cells. This indicates that the germline stem cells directly control fat storage.

Using RNA interference in screens that reduce gene expression, Wang *et al.* identified a triglyceride lipase that, when eliminated, reversed the reduction of fat storage and longevity caused by germline stem cell depletion. Overexpression of the lipase reduced fat storage and extended life span. Further analysis indicated that the lipase regulates fat storage in intestinal cells. These findings indicate that fat metabolism is the missing link between the status of the reproductive system and longevity in the worm. Thus, in response to germline stem cell depletion, the authors propose that somatic cells of the gonad send a signal to reduce fat storage, thereby extending life span.

Previous studies in the worm have also shown that the effects of both the IIS pathway and germline stem cells on longevity require FOXO/DAF-16 (*9*, *11*). As expected, Wang *et al.* observed that disrupting the IIS pathway increased lipase expression in intestinal cells, reduced fat storage, and thus increased worm life span. By contrast, reducing lipase expression partially suppressed longevity caused by a defective IIS pathway. In addition, inactivating FOXO/DAF-16 did not decrease lipase expression nor increase fat storage, but it restored fat storage caused by germline stem cell depletion. Furthermore, although *kri-1* and *daf-12* are important in intestinal cells for activating FOXO/DAF-16 in response to germline stem cell depletion, *kri-1* is required for, whereas *daf-12* is dispensable for, increasing lipase expression and reducing fat storage. These findings indicate that germline stem cell depletion promotes lipid metabolism in intestinal cells by activating a KRI-1–DAF-16 signaling pathway but not the hormone pathway that involves DAF-12. Therefore, the IIS pathway and reproductive signaling converge on FOXO/DAF-16 to control fat metabolism and longevity.

A recent genetic study in *C. elegans* indicates that under the condition of reducing signaling from the IIS pathway below a critical threshold level, the somatic gonad is no longer needed for germline stem cell depletion to extend life span, which suggests that the signals from the reproductive system somehow also impinge on the IIS pathway (*15*). The findings of Wang *et al.* and other studies support a model in which somatic cells of the gonad send signals that impinge on intestinal cells and control two pathways—the KRI-1–DAF-16 pathway and the DAF-12 pathway—to increase longevity (see the figure). The IIS pathway also functions through DAF-16 to control fat metabolism and life span. Signals from germline stem cells block these pathways and shorten life span.

In *C. elegans*, it will be critical to identify the life span–extending signals produced by the reproductive system that act on intestinal cells to control longevity, and to elucidate the molecular details of their effect. Germline stem cell depletion also extends life span in *Drosophila* (*6*), so it is of interest to investigate whether this also involves fat metabolism. Finally, it will be exciting to determine whether germline stem cell depletion can extend life span by regulating fat metabolism in mammals. If fat metabolism is the conserved link between reproduction and aging, we may discover more about how life span is controlled in humans and perhaps find better treatments for age-related diseases.

References

1. J. Vijg, J. Campisi, *Nature* **454**, 1065 (2008).
2. M. D. Piper, A. Bartke, *Cell Metab.* **8**, 99 (2008).
3. H. Hsin, C. Kenyon, *Nature* **399**, 362 (1999).
4. C. M. Sgrò, L. Partridge, *Science* **286**, 2521 (1999).
5. N. Arantes-Oliveira, J. Apfeld, A. Dillin, C. Kenyon, *Science* **295**, 502 (2002).
6. T. Flatt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6368 (2008).
7. M. C. Wang, E. J. O'Rourke, G. Ruvkun, *Science* **322**, 957 (2008).
8. M. Tatar, A. Bartke, A. Antebi, *Science* **299**, 1346 (2003).
9. K. Lin, H. Hsin, N. Libina, C. Kenyon, *Nat. Genet.* **28**, 139 (2001).
10. A. Dillin, D. K. Crawford, C. Kenyon, *Science* **298**, 830 (2002).
11. J. R. Berman, C. Kenyon, *Cell* **124**, 1055 (2006).
12. M. E. Giannakou *et al.*, *Science* **305**, 361 (2004); published online 10 June 2004 (10.1126/science.1098219).
13. D. S. Hwangbo, B. Gershman, M. P. Tu, M. Palmer, M. Tatar, *Nature* **429**, 562 (2004).
14. M. Blüher, B. B. Kahn, C. R. Kahn, *Science* **299**, 572 (2003).
15. T. M. Yamawaki *et al.*, *Genetics* **178**, 513 (2008).

10.1126/science.1166150

GEOPHYSICS

Reconstructing Earth History in Three Dimensions

Bernhard Steinberger

An inverse model elucidates the connection between plate tectonics at Earth's surface and the dynamics of the underlying mantle.

Plate motions at Earth's surface are intimately linked to convective flow in the underlying mantle. These links are becoming more evident through subsurface tomographic images, advances in mineral physics, and improved models of plate motion. Yet, there is still no generally accepted mechanism that consistently explains plate tectonics in the framework of mantle convection. A key obstacle to a better understanding is the fact that, although the configuration of plates at Earth's surface can be reconstructed with some confidence at least back to the Cretaceous, knowledge about the deep interior has been largely limited to the present day. On page 934 of this issue, Liu *et al.* (*1*) point the way to more reliable reconstructions of Earth's past in three dimensions.

The theory of plate tectonics describes how plates move away from each other at spreading ridges, past each other at transform faults, and toward each other at convergent boundaries. In the latter case, one of the plates may get subducted back into Earth's interior. Seismic tomography, which uses travel times between earthquakes and seismographs to construct three-dimensional models of seismic wavespeed distribution of Earth's interior, has yielded images of such subducted "slabs" in ever-increasing detail (*2*). Using subduction inferred from plate reconstructions as input, present-day mantle

The Center for Geodynamics, Geological Survey of Norway, 7491 Trondheim, Norway. E-mail: bernhard.steinberger@ngu.no

temperature structure can be numerically forward-modeled and the model results compared with the inferred temperature structure based on seismic tomography and mineral physics. By modifying model assumptions, the agreement can be optimized. In this way, forward models have yielded important insights into plate reconstructions and mantle properties (*3*, *4*).

Yet, an obvious drawback of this approach is that it cannot achieve an exact fit between predicted and observed present-day mantle structure. Hence, an alternative approach has been to begin with present-day structure inferred from tomography and compute mantle structure backward in time (*5*). But past mantle structure cannot be completely recovered in this way because information has been lost due to thermal diffusion. At least in the thermal boundary layers at the top and bottom of the mantle, heat transport through thermal diffusion cannot be neglected. However, time-reversing thermal diffusion makes—backward in time—hot material become hotter and cold material colder. Eventually, this effect is bound to create numerical instabilities.

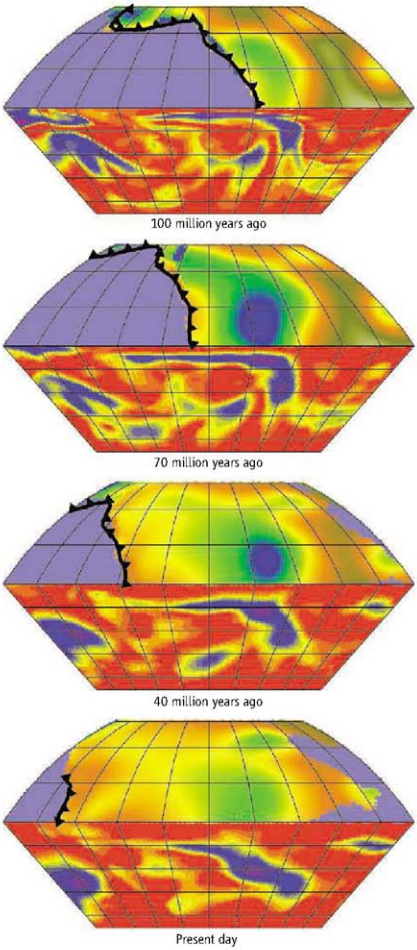
To overcome this problem, inverse approaches to mantle convection have been developed (*6*, *7*), which essentially aim at finding the initial condition, starting from which the forward model optimally recovers the present-day state of the mantle. Solving this problem requires multiple iterations and is computationally intensive. Liu *et al.* now use this approach to obtain a realistic reconstruction of past mantle structure (see the figure).

Plate reconstructions (*8*) provide the surface boundary condition of mantle flow. However, that alone does not guarantee that the reconstruction is geologically reasonable. This is in fact one main problem that Liu *et al.* face: Without modification, the past location of slabs is reconstructed too far east to be compatible with plate reconstructions. The authors overcome this problem by introducing a "stress guide," which essentially allows the Farallon plate (the oceanic plate subducted under the west coast of North America during the past 100 million years) (see the figure) to stuff material sideways under North America. However, it remains unclear how, in the real Earth, such a stress guide could be accomplished, and more generally, how models of the past mantle can be made compatible with plate tectonic reconstructions.

One important aspect of this goal is to devise a common reference frame for plate motions and mantle dynamics (*9*), in particular to distinguish between plate motions relative to the mantle and coherent rotations of the entire Earth relative to its spin axis (true polar wander). Through optimizing agreement between plate tectonic reconstructions of subduction and mantle dynamic reconstructions of slabs, models such as that of Liu *et al.* will be instrumental in finding a common reference frame.

What is all this modeling useful for? Liu *et al.* show one main application: predicting vertical surface motions. Dense material in the mantle causes downward flow and thus pulls down Earth's surface. Hence, when a continent moves over the remnants of a subducted slab, parts of it may get flooded (*10*). Knowing when and where continental flooding has occurred is not merely of academic interest, because the flooding history also influences where sediments and related natural resources may form. Knowledge of uplift and subsidence is also necessary for extracting information about past sea-level change from the geologic record (*5*) and understanding which part of it is caused by ocean basin volume change. It is thus essential for understanding past climates and ultimately helpful for better understanding causes and effects of present climate change.

Many challenges remain. Liu *et al.* avoid unnecessary complexity, using simple models of



Going down. Mantle temperature is shown at the front, continental dynamic topography at the top. Ocean floor is subducted along a trench (black line with triangles) at the western edge of North America. The subducted slab (in "cold" colors on the temperature cross section) causes negative topography (green and blue colors on top). The North American plate overrides this topography low, which becomes less pronounced as the slab sinks toward the base of the mantle. All data provided by L. Liu (*1*). The figure was prepared using the GMT (*11*) software.

mantle rheology for relating seismic wavespeed variations to temperature. Future studies will aim at making models gradually more Earth-like. A key aspect of this will be to develop more realistic models of lithospheric strength, which controls the amount of deflection that can develop due to buoyant mantle loads, and drag of the flowing mantle at the base of the lithosphere, which can lead to lithospheric deformation. Improved models of topography development must link mantle dynamics to these more realistic lithosphere models. Toward this goal, the work of Liu *et al.* is a major step in linking plate tectonics and mantle dynamics.

References and Notes

1. L. Liu, S. Spasojević, M. Gurnis, *Science* **322**, 934 (2008).
2. H. Bijwaard, W. W. Spakman, E. R. Engdahl, *J. Geophys. Res.* **103**, 30,055 (1998).
3. H.-P. Bunge, S. P. Grand, *Nature* **405**, 337 (2000).
4. E. Hafkenscheid *et al.*, *Tectonophysics* **333**, 35 (2001).
5. R. D. Müller *et al.*, *Science* **319**, 1357 (2008).
6. H.-P. Bunge, C. R. Hagelberg, B. J. Travis, *Geophys. J. Int.* **152**, 280 (2003).
7. A. Ismail-Zadeh, G. Schubert, I. Tsepelev, A. Korotkii, *Phys. Earth Planet. Inter.* **145**, 99 (2004).
8. See www.gplates.org.
9. T. H. Torsvik *et al.*, *Rev. Geophys.* **46**, RG3004, 10.1029/2007RG000227 (2008).
10. M. Gurnis, *Nature* **344**, 754 (1990).
11. P. Wessel, W. H. F. Smith, *Eos* **72**, 441 (1991).

12. The author thanks T. Torsvik for help with the manuscript. The work is funded by the Geological Survey of Norway.

10.1126/science.1166395

APPLIED PHYSICS

Plasmonics Applied

Albert Polman

Surface plasmons, light-induced excitations of electrons on metal surfaces, may provide integration of electronics and optics on the nanoscale.

The ability to engineer metal surfaces and particles at the nanoscale has led to the rapid development of the field of "plasmonics," the optical properties of metal structures at the nanoscale. Surface plasmons are optically induced oscillations of the free electrons at the surface of a metal. Electron beam lithography, focused ion beam milling, and self-assembly have provided routes to engineer complex arrays of metal nanostructures in which plasmons can be excited, directed, and manipulated. The attractiveness of plasmons is that they can effectively confine the optical excitation in a nanoscale volume and thus mediate strong optical interactions within this volume. Also, the wavelength at which these phenomena are observed can be tuned by varying the metal shape and dielectric environment, thereby providing a broad palette from which to choose the desired optical properties for an application.

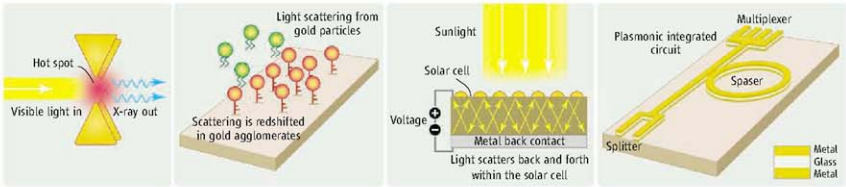
Early work in plasmonics focused on the study of resonances and electromagnetic field enhancements in individual metal nanoparticles and particle assemblies. The plasmon coupling within arrays of metal nanoparticles can lead to the formation of nanoscale hot spots in which the intensity of light from an incident beam can be concentrated by more than four orders of magnitude. This leads to a large improvement in sensing techniques that use optical radiation, such as Raman spectroscopy, with potential applications in medical diagnostics. The effect of light concentration via plasmons is most apparent in phenomena that are nonlinear in light intensity, as demonstrated recently by the on-chip generation of extreme-ultraviolet light by pulsed-laser high harmonic generation (*1*). This opens a wealth of prospects in lithography or imaging at the nanoscale through the use of soft x-rays (see the figure, left panel).

Because the plasmonic interaction between metal nanoparticles is very sensitive to their separation, precise measurements of the plasmon resonance wavelength of metal particle assemblies functionalized with biomolecules can be used as a molecular-scale ruler that operates over a length scale much larger than that in the fluorescence energy transfer metrology that is routinely used in biology (*2*). Practical applications of this concept in systems biology, such as imaging of the motion of molecular motors, are being pursued. Already, measurement of plasmonic resonance shifts is used in the detection of biomolecules (see the figure, middle left panel), and indeed standard commercial pregnancy tests are based on this principle. A potentially far-reaching application is the use of particles composed of a dielectric core and a metallic shell (*3*) in cancer treatment. These particles, when injected into the human body, are selectively bound to malicious cells, whereupon laser irradiation at a precisely engineered plasmon resonance wavelength is used to heat the particles and thereby destroy the cells. Clinical studies are showing promising results (*4*).

Suitably engineered metal nanostructures can also act as antennas in which the resonant coupling between the particles concentrates light into well-defined hot spots (*5*), enabling ultrasmall, wavelength-sensitive directional sensors or detectors. The same metal particle arrays, when coupled to optical emitters, can also act as directional emitters. Indeed, the enhanced optical density of states near the surface of metal nanoparticles can provide control over the color, directionality, and polarization of light-emitting diodes. This concept may find large-scale applications in the areas of solid-state lighting and photovoltaics (see the figure, middle right panel). Calculations and experiments (*6*) show that light scattering from metal nanoparticle arrays placed on top of a thin-film semiconductor layer can effectively fold the path of sunlight into the layer, strongly enhancing its effective absorption.

Parallel to the development of plasmonic structures based on metal nanoparticles, the propagation of plasmons along metal waveguides is also being investigated. Here too, precise control over material and geometry allows the wave-guiding properties to be controlled in ways that cannot be achieved with regular dielectric waveguides. In particular, extremely short wavelengths can be achieved at optical frequencies. It has been shown that light with a free-space wavelength of 651 nm, squeezed in a metal-insulator-metal plasmonic waveguide, has its wavelength shrunk to only 58 nm (*7*). The next challenge will be to shrink it further, into the soft x-ray wavelength regime. Similar to the coupling within nanoparticle assemblies, this effect is due to the coupling between plasmons propagating at the two metal-insulator interfaces. By further tailoring plasmonic waveguide structures, the propagation speed of plasmons can be reduced well below the speed of light (*8*). More complex geometries, in which arrays of nanoholes are integrated in a metal film, act as efficient color filters (*9*). Interestingly, in some geometries, plasmon waveguides exhibit a negative refractive index for the guided plasmon, and indeed two-dimensional negative refraction has been observed in these plasmonic waveguides (*10*).

Center for Nanophotonics, FOM-Institute AMOLF, Kruislaan 407, 1098 SJ Amsterdam, Netherlands. E-mail: polman@amolf.nl



Applied plasmonics. (Left) A plasmonic hot spot between metal nanoparticles creates soft x-rays. **(Middle left)** Measuring the resonance shift in coupled metal nanoparticles leads to efficient sensing. **(Middle right)** Scattering from metal nanoparticles enhances light trapping in a solar cell. **(Right)** Plasmonic integrated circuit with subwavelength dimensions. A plasmonic ring laser is integrated with 50-nm-wide waveguides.

These studies on planar plasmon propagation will lead to the design of plasmonic integrated circuits (see the figure, right panel) in which optical information is generated, switched, amplified, and detected within dimensions much smaller than the (free-space) wavelength of light. This will enable the integration of optics with nanoscale semiconductor integrated circuit technology, which so far has proven impossible because of the different length scales of optics and electronics. Although plasmons decay as they propagate, it should be realized that the relatively small propagation lengths can be tolerated in devices with nanoscale dimensions. The eagerly awaited demonstration of a plasmon laser or amplifier would be a further enabler of this plasmo-electronic technology.

A recent development in plasmonics is the study of metallic nanoresonators in two- and three-dimensional arrangements, to form optical "metamaterials" with artificially engineered permittivity and permeability. Negative refraction of near-infrared light has been demonstrated in a stack of metallic "fishnet" structures (*11*). Because of the peculiar way light is refracted in a material with negative index, it is possible to achieve subwavelength optical imaging (*12*). Suitably engineered negative-index geometries may even act as invisibility cloaks for wavelengths in the visible regime.

In only a few years, plasmonics has grown from a field in which fundamental insights were developed to an area that demonstrates important applications. Many new fundamental research topics are currently being pursued, including attosecond dynamics and coherent control of plasmons (*13*), lasing spasers (*14*), cloaking using novel geometries (*15*), and quantum mechanical effects at the subnanoscale level. These studies, which are very exploratory and rich in new science, will very likely lead to new exciting applications of plasmonics as well.

References

1. S. Kim *et al.*, *Nature* **453**, 757 (2008).
2. C. Sonnichsen *et al.*, *Nat. Biotechnol.* **23**, 741 (2005).
3. E. Prodan, C. Radloff, N. J. Halas, P. Nordlander, *Science* **302**, 419 (2003).
4. C. Loo *et al.*, *Nano Lett.* **5**, 709 (2005).
5. R. de Waele, A. F. Koenderink, A. Polman, *Nano Lett.* **7**, 2004 (2007).
6. S. Pillai, K. R. Catchpole, T. Trupke, M. A. Green, *J. Appl. Phys.* **101**, 093105 (2007).
7. H. T. Miyazaki, Y. Kurokawa, *Phys. Rev. Lett.* **96**, 097401 (2006).
8. M. Sandtke, L. Kuipers, *Nat. Photon.* **1**, 573 (2007).
9. E. Laux, C. Genet, T. Skauli, T. W. Ebbesen, *Nat. Photon.* **2**, 161 (2008).
10. H. J. Lezec, J. A. Dionne, H. A. Atwater, *Science* **316**, 430 (2007); published online 21 March 2007 (10.1126/science.1139266).
11. J. Valentine *et al.*, *Nature* **455**, 376 (2008).
12. J. B. Pendry, *Phys. Rev. Lett.* **85**, 3966 (2000).
13. M. I. Stockman, M. F. Kling, U. Kleineberg, F. Krausz, *Nat. Photon.* **1**, 539 (2007).
14. N. I. Zheludev *et al.*, *Nat. Photon.* **2**, 351 (2008).
15. A. Alu, N. Engheta, *Phys. Rev. Lett.* **100**, 113901 (2008).

10.1126/science.1163959

NEUROSCIENCE

Overcoming Inhibitions

Two reasons why injured neurons of the central nervous system fail to regenerate are inhibition by myelin proteins and a signaling pathway that blocks axon growth.

Woo-Yang Kim and William D. Snider

Nervous system injuries and neurological diseases often result in loss of sensation and paralysis that are irreversible. This is in part because neurons in the central nervous system are unable to regrow long cellular processes (axons) after they are damaged. Thus, neurons lose the interconnections that are vital for nervous system function. On page 967 and 963 of this issue, Atwal *et al.* (*1*) and Park *et al.* (*2*) reveal two molecular mechanisms that contribute to this failure to regenerate, pointing to new potential therapeutics.

Until the early 1980s, there was little progress in understanding axon regeneration. One of the first breakthroughs came from Aguayo and colleagues, who showed that severed axons of mammalian central nervous system neurons could be induced to grow when placed next to a transplanted peripheral nerve, which was presumed to provide a favorable growth environment (*3*). Subsequently, Schwab and colleagues showed that glial cells (oligodendrocytes) that envelop axons with myelin inhibit axon growth, and that an antibody against myelin could reverse this effect (*4*). The glial scar that forms at the site of injury is also known to be highly inhibitory (*5*). Finally, we now know that even though axons can be induced to grow in a favorable environment, intrinsic developmental changes limit their growth capacity. Compared to neurons in the peripheral nervous system, those in the central nervous system drastically reduce their growth capacity during development (*6*).

Several groups have characterized specific components of myelin that block axon growth, including myelin-associated glycoprotein (MAG), Nogo-A, and oligodendrocyte myelin glycoprotein (OMgp) (*7*). A receptor for Nogo-A (NgR), expressed on neurons, was identified (*8*) and found to also bind MAG and OMgp (*7*). However, the

Neuroscience Center, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: william_snider@med.unc.edu

mechanisms by which these proteins block axon growth appear complex. The neurotrophin receptor p75, a tumor necrosis factor receptor superfamily member, interacts with NgR and is required for the inhibitory effects of NgR in vitro (*9*, *10*). The identification of p75 suggested a plausible link between the inhibition of regeneration and the neuronal actin cytoskeleton because p75 can activate the small guanosine triphosphatase RhoA, a regulator of the actin dynamics. Lingo-1 is yet another co-receptor that is involved in Rho activation by the p75-NgR complex (*11*). Finally, TROY, another tumor necrosis factor receptor superfamily member, forms a complex with NgR and Lingo-1 and is thought to mediate the inhibitory effects of myelin in neurons that lack p75 (*12*, *13*). Although MAG, Nogo, and OMgp inhibit axon growth in vitro, mutant mice that lack these proteins, as well as mice that lack NgR, exhibit only modest axon regeneration and functional recovery after injury (*14*). Additional inhibitory components in myelin, and their cognate receptors on neurons, are thus thought to be involved in controlling axon regeneration.

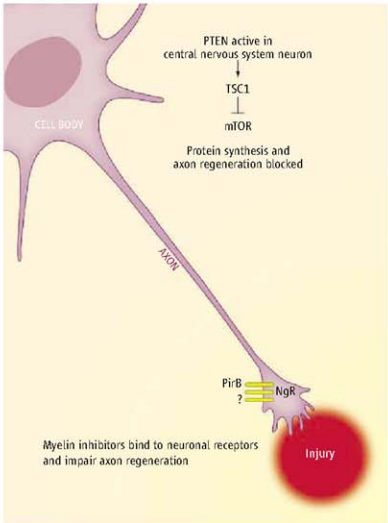
Atwal *et al.* indeed identify another receptor for myelin-inhibitory proteins, leukocyte immunoglobulin–like receptor B2 (LILRB2), which has five homologous family members in humans. Mice have a single ortholog of LILRB2 called Paired immunoglobulin-like receptor B (PirB), which is also a receptor for myelin-inhibitory proteins (see the figure). Genetic ablation of PirB or inhibition of PirB with a specific antibody partially blocked the effects of myelin inhibitors on axon growth in cultured mouse cerebellar granule neurons. Interestingly, PirB binds to SHP phosphatases in neurons (*15*), enzymes involved in modulating the cytoskeleton, suggesting a possible signaling pathway that blocks axon growth.

Inhibiting PirB in NgR-deficient cerebellar granule neurons almost completely rescued axon growth on myelin extracts, but only partially rescued the axon growth inhibition by Nogo. The former effect demonstrates that PirB and NgR function cooperatively to mediate the effects of myelin, whereas the latter result indicates that unknown additional receptors may transmit Nogo's inhibitory effects in certain types of neurons.

Both PirB and NgR define a critical period during mammalian development when marked plasticity is evident in the visual system (*15*, *16*). Myelin-inhibitory proteins may thus help constrain the short-range axon growth and sprouting that underlie the continual rearrangement of neuron-neuron connections (synapses). Indeed, the evidence that myelin-inhibitory proteins are involved in neural plasticity is stronger than the evidence that they block long-range axon regeneration after injury.

Despite considerable efforts to increase the intrinsic regenerative capability of mature central nervous system neurons, treatment with neuronal growth factors produces only modest effects in most experimental paradigms. Park *et al.* provide an important advance related to the reduction in intrinsic growth potential of central nervous system neurons. The authors reasoned that the molecular machinery underlying the growth and proliferative capacity of tumors might be harnessed to enhance growth of central nervous system neurons, and that the actions of tumor suppressor genes might constrain regenerative capacity in mature neurons. To test these ideas, Park *et al.* exploited the optic nerve crush paradigm in mice to assess regeneration efficacy. This paradigm is appealing because normally very few axons of retinal ganglion cells cross the site of injury (crush site), and most ganglion cells die within 2 to 3 weeks after the procedure.

Park *et al.* used a conditional gene "knockout" approach in which a particular tumor suppressor gene was targeted for deletion by injecting a specific agent into the retina. In this case, the agent was a virus that expressed a DNA recombinase (Cre), which can remove a specified gene in a mouse if the animal has been genetically engineered to express Cre binding sites that flank the targeted gene (the Cre-Lox system). The retina is ideal for this technique because virus can be directly injected into the eye and infect almost all retinal ganglion cells. Park *et al.* found that eliminating the gene encoding the tumor suppressor p53 enhanced the survival of retinal ganglion cells, but not axon regeneration. However, eliminating the gene encoding the tumor suppressor PTEN not only prevented apoptosis (programmed cell death) of retinal ganglion cells, but also promoted robust axon extension. Approximately 10% of the injured axons regenerated, and some retinal ganglion cells extended very long axons. This regrowth is comparable to that of retinal ganglion cell axon regeneration after optic nerve crush in rats induced by lens injury (which stimulates release of a novel growth factor, oncomodulin) plus administration of a Rho inhibitor (*17*, *18*). Given that mature retinal ganglion cells lose intrinsic axon growth capacity and that the mature optic nerve is a strongly inhibitory environment, the robust axon regeneration of retinal ganglion cells observed by eliminating PTEN alone is surprising and exciting.



Regeneration in the central nervous system. When PTEN is present, protein synthesis that is stimulated by the mTOR signaling pathway is blocked. PTEN may also elicit growth-promoting effects through other pathways, such as the one that controls GSK-3. PirB, NgR, and other receptors yet to be identified mediate the inhibitory effects of myelin proteins on axon regeneration.

Park *et al.* further determined that axon regeneration induced by PTEN deletion is partly mediated by the mTOR signaling pathway (see the figure). This pathway controls protein synthesis by regulating the activity of ribosomal S6 kinase and eukaryotic initiation factor 4E binding protein. The authors found that mTOR signaling decreased in most retinal ganglion cells during development. Treatment of retinal ganglion cells with rapamycin, an inhibitor of mTOR, abolished

the regenerative effect of PTEN deficiency. Genetic ablation of a negative mTOR regulator, tuberous sclerosis complex 1 (TSC1), partially but not completely mimicked the effects of PTEN deficiency on axon regeneration, indicating that other PTEN-regulated pathways such as glycogen synthesis kinase–3 (GSK-3) could be involved in controlling axon growth.

It is uncertain whether the intrinsic regenerative mechanism observed by Park *et al.* is vigorous enough to overcome a hostile extrinsic environment. For example, an optic nerve crush produces much less inflammation and glial scarring than a contusion injury that damages the spinal cord. It will be important to determine the extent to which the findings of Park *et al.* generalize to other neuronal populations such as corticospinal axons and to understand why mTOR activity is reduced during development and after axon injury.

Will the studies by Atwal *et al.* and Park *et al.* lead to advances in treating human spinal cord injuries? Prior work with myelin-inhibitory proteins has initiated exploratory clinical trials. The identification of LILRB2 as a human receptor for myelin-inhibitory proteins should stimulate new thinking in this area. However, work with primates, which more adequately model human injuries than rodents, is just beginning (*19*). It is unclear whether therapeutic approaches centered around PTEN inhibition could be developed, and whether PTEN inhibitors can mimic the positive effects of deleting the PTEN gene on axon regeneration. Nevertheless, the idea of enhancing protein synthesis to promote long-distance axon growth after injury is an appealing possibility.

References

1. J. K. Atwal *et al.*, *Science* **322**, 967 (2008).
2. K. K. S. Park *et al.*, *Science* **322**, 963 (2008).
3. P. M. Richardson *et al.*, *Nature* **284**, 264 (1980).
4. P. Caroni, M. E. Schwab, *Neuron* **1**, 85 (1988).
5. J. Silver, J. H. Miller, *Nat. Rev. Neurosci.* **5**, 146 (2004).
6. J. L. Goldberg *et al.*, *Science* **296**, 1860 (2002).
7. G. Yiu, Z. He, *Nat. Rev. Neurosci.* **7**, 617 (2006).
8. A. E. Fournier *et al.*, *Nature* **409**, 341 (2001).
9. K. C. Wang *et al.*, *Nature* **420**, 74 (2002).
10. S. T. Wong *et al.*, *Nat. Neurosci.* **5**, 1302 (2002).
11. S. Mi *et al.*, *Nat. Neurosci.* **7**, 221 (2004).
12. Z. Shao *et al.*, *Neuron* **45**, 353 (2005).
13. J. B. Park *et al.*, *Neuron* **45**, 345 (2005).
14. B. Zheng *et al.*, *Trends Neurosci.* **29**, 640 (2006).
15. J. Syken *et al.*, *Science* **313**, 1795 (2006).
16. A. W. McGee *et al.*, *Science* **309**, 2222 (2005).
17. D. Fischer *et al.*, *J. Neurosci.* **24**, 8726 (2004).
18. Y. Yin *et al.*, *Nat. Neurosci.* **9**, 843 (2006).
19. S. Rossignol *et al.*, *J. Neurosci.* **27**, 11782 (2007).

10.1126/science.1166152

CELL BIOLOGY

Going Global on Ubiquitin

A new technique that profiles protein stability provides a powerful platform in which high-throughput screening can be performed in real time with single-cell resolution.

Caroline Grabbe[1] and Ivan Dikic[1,2,3]

Our vision of protein control has for many years been viewed from a transcriptional or activity-based perspective. More recently, protein stability and regulated degradation have emerged as equally important issues to address. On pages 918 and 923 of this issue, Elledge and colleagues describe a new technology to analyze global protein stability (GPS). The studies introduce the new approach (*1*) and illustrate how it can be applied to identify substrates of a specific enzyme (ubiquitin ligase) (*2*).

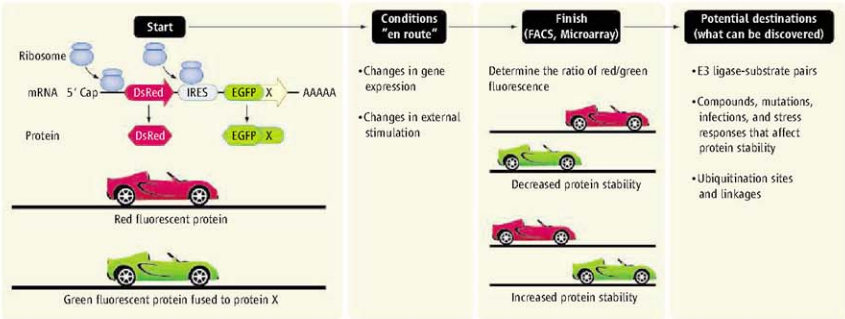
In addition to protein degradation that occurs in the lysosomal compartment of cells, degradation by a cellular shredding machine known as the proteasome is another major route to eliminate proteins. Targeting to the proteasome is preceded by the addition of ubiquitin chains to the selected substrates, an event catalyzed by the sequential action of activating, conjugating, and ligating enzymes (E1, E2, and E3, respectively). The specificity of substrate recognition is mediated mainly by divergent use of the estimated ~617 ubiquitin ligases encoded by the human genome (*3*). E3 ligases fall into different classes, based on structural composition and mechanism of action. In some cases, a group of proteins comes together to form multisubunit ubiquitin ligases, as is the case for the SCF (Skp1-cullin-F-box) complex where Skp1, Cul1, Rbx1, and an F-box protein form the core of the ligase (*4*).

Despite extensive efforts to map substrate targeting by individual ligases, the methods used have been laborious and the results far from complete. Most studies have used direct substrate-ligase interaction as a basis for substrate identification and have thus been biased toward strongly interacting targets (*5*). The GPS approach offers a new mode to navigate the ubiquitin-proteasome system and identify substrates for a given E3 ligase or, in general, to investigate how a chemical or physical stimulus affects the stability of a given protein. The system evaluates protein abundance at a global level in living cells, with accuracy comparable to conventional time-consuming experiments that analyze only a few proteins at a time. Eight thousand distinct complementary DNAs were used to generate a library of cultured human cells in which each cell expresses a common stable red fluorescent protein (DsRed) together with a variable fusion protein composed of enhanced green fluorescent protein (EGFP) and a unique open reading frame (ORF), produced from the same transcript.

The turnover of the EGFP-ORF fusion proteins can thus be monitored by flow cytometry (which counts, examines, and sorts whole cells) as a ratio of red to green fluorescence in each cell (see the figure). Cells in which the fluorescence ratio changes in response to a gene perturbation or stimulus can be sorted by the degree of change, and the identity of the ORFs they express can be easily identified by a polymerase chain reaction–based microarray approach. During the development of GPS profiling, a comparative analysis of all ORFs tested was used to assign each ORF a protein stability index value that roughly categorizes each corresponding protein as having a short, medium, long, or extra-long life span. An impressive power of the GPS method is the capacity of single-cell resolution, which is in contrast to other established methods that frequently generate population-averaged readouts. In addition, measurements can take place in live cells, in real time, and can be integrated with systems for automation to enable high-throughput studies.

The GPS approach is a major advance in the quest to gain a comprehensive understanding of protein turnover in cells and will be a valuable complement to the biophysical methods that have emerged in the past 5 years to analyze substrate ubiquitination (*6*). Large-

[1]Institute of Biochemistry II and Cluster of Excellence Macromolecular Complexes, Goethe University, Frankfurt am Main, Theodor-Stern-Kai 7, D-60590 Frankfurt (Main), Germany. [2]Mediterranean Institute for Life Sciences, 21000 Split, Croatia. [3]Department of Immunology, School of Medicine, University of Split, Soltanska 2, 21000 Split, Croatia. E-mail: ivan.dikic@biochem2.de



Navigating the world of ubiquitin and protein degradation. In GPS profiling, the ratio between a constant factor (DsRed) and a variable factor (EGFP-X) is measured by a combinatorial approach of flow cytometry (FACS analysis) and microarray technology. This serves as a readout of protein abundance and stability. By challenging the system with internal modifications or external stimuli, the responses can be monitored on a global scale. Many questions concerning the process of protein degradation, including E3 ligase specificity, routes to destruction, and importance of ubiquitin linkage, can be addressed.

scale analysis of ubiquitinated proteins has relied mainly on a combination of affinity purification and mass spectrometry analysis of proteins in yeast (*7*), human cell lines (*8*), and transgenic mice (*9*). To more specifically analyze ubiquitinated targets downstream of individual E3 ligases, Ota *et al.* used SILAC (stable isotope labeling with amino acids in cell culture) to quantify the overall change in protein ubiquitination after alterating E3 ligase activity (*10*). Substrates of the E3 ligase Rsp5 were recently identified in a high-throughput assay in which all proteins expressed in yeast were spotted onto a nitrocellulose chip and directly tested for ubiquitination by Rsp5 in vitro (*11*).

The GPS technology moves the field closer toward global in vivo mapping of ligase-substrate pairs. Illustrating the feasibility of the system, Yen and Elledge used GPS profiling to identify substrates of the SCF ubiquitin ligase complex (*2*). Broadly outlined, in the background of the described GPS library, SCF function was abrogated by the expression of a dominant negative Cul1, whereupon the perturbed cells were analyzed by GPS profiling. An impressive number of 359 targets were postulated as putative SCF substrates, among which 66 were tested and 31 verified. In a majority of cases, verification of a specific protein was accomplished by analyzing individual samples by flow cytometry [fluorescence-activated cell sorting (FACS) analysis] as well as by biochemical detection in cell extracts with antibodies (immunoblotting). In the future, combining a GPS-based ORF library with methods that perturb gene expression on a global scale, such as genome-wide RNA interference, leaves few limitations to the amount of knowledge that may be acquired with this innovative method.

A feature that is both a strength and a limitation is that the GPS technique does not monitor E3 substrates directly but rather the outcome of E3 activity. Thus, there is no discrimination between the direct and indirect effects of an E3 ligase. However, this could be resolved by following the kinetics of protein degradation. GPS profiling is also biased toward identifying ubiquitinated substrates that are destined for either degradation or stabilization. This excludes proteins that are functionally affected by the modification with regard to their enzymatic activity, localization in the cell, and ability to form complexes with other cellular constituents. Nevertheless, together with conventional proteomic approaches, the GPS system will provide a powerful means to distinguish the consequences of different types of ubiquitination, sorting the proteolytic events from those of regulatory nature (*12*).

In addition to assigning ligase-substrate pairs, GPS profiling has the potential to elucidate the essence of differential ubiquitin chain linkages, in particular how linkage affects the efficiency of proteasomal targeting. It may also help to identify degradation signals (degrons) encoded by amino acid sequences, to associate specific lysine residues that are modified by ubiquitin to functionality, and to explain why some proteins are directly routed to the proteasome whereas others require shuttle factors to ensure proper targeting. There is also the potential to transfer GPS profiling into model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*, creating ORF libraries that will enable substrate screening in vivo.

Given the importance of the ubiquitin-proteasome system for cellular functions, dysfunctions of the involved players clearly have the capacity to cause disease. Indeed, defective ubiquitination and protein degradation is implicated in the etiology of cancer and neurodegenerative disorders, among others (*13*, *14*). In this context, the GPS method could be used to screen for compounds that counteract such deficiencies. Another interesting aspect would be to investigate how protein degradation is altered when cells are exposed to infectious agents or various stress situations. Overall, the GPS approach will likely become an important component of the integrated approaches needed to systematically map the mechanisms of regulated protein degradation.

References

1. H.-C. S. Yen, Q. Xu, D. M. Chou, Z. Zhao, S. J. Elledge, *Science* **322**, 918 (2008).
2. H.-C. S. Yen, S. J. Elledge, *Science* **322**, 923 (2008).
3. W. Li *et al.*, *PLoS ONE* **3**, e1487 (2008).
4. T. Ravid, M. Hochstrasser, *Nat. Rev. Mol. Cell Biol.* **9**, 679 (2008).
5. A. Peschiaroli *et al.*, *Mol. Cell* **23**, 319 (2006).
6. J. Peng, *BMB Rep.* **41**, 177 (2008).
7. J. Peng *et al.*, *Nat. Biotechnol.* **21**, 921 (2003).
8. M. Matsumoto *et al.*, *Proteomics* **5**, 4145 (2005).
9. H. B. Jeon *et al.*, *Biochem. Biophys. Res. Commun.* **357**, 731 (2007).
10. K. Ota, K. Kito, S. Okada, T. Ito, *Genes Cells* **13**, 1075 (2008).
11. R. Gupta *et al.*, *Mol. Syst. Biol.* **3**, 116 (2007).
12. F. Ikeda, I. Dikic, *EMBO Rep.* **9**, 536 (2008).
13. G. Nalepa, M. Rolfe, J. W. Harper, *Nat. Rev. Drug Discov.* **5**, 596 (2006).
14. D. Hoeller, C. M. Hecker, I. Dikic, *Nat. Rev. Cancer* **6**, 776 (2006).

10.1126/science.1166845

EPPENDORF WINNER

Switching Memories ON and OFF

Control of new protein synthesis regulates whether an experience is translated into a long-term memory.

Mauro Costa-Mattioli

Ireneo Funes, the fictional main character in Jorge Luis Borges' short story "Funes el Memorioso," could remember in vivid detail every day of his life after he was thrown from a wild horse at a ranch in Fray Bentos, Uruguay. He had acquired a prodigious ability to store new information without any practice. Unlike Funes [and real "autistic savants" (*1*)], who could store information with a glance, most people learn new things only after many attempts.

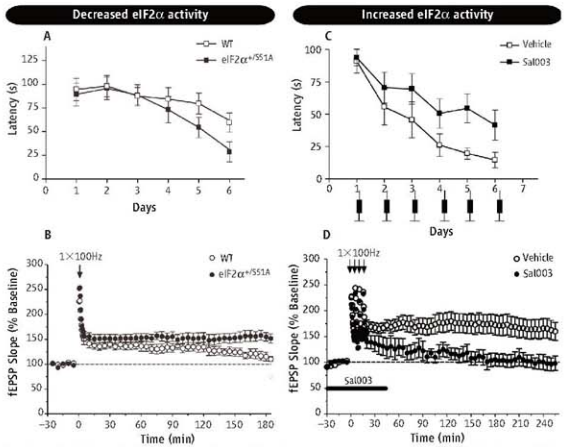
Psychologists have identified two corresponding processes: short-term memory, which lasts from seconds to minutes; and long-term memory, which lasts for days, months, or even a lifetime (*2*, *3*). It is now well accepted that making long-lasting memories is dependent on the ability of brain cells (neurons) to synthesize new proteins (*2*–*7*). Indeed, animals treated with a drug that blocks the production of new proteins cannot form long-term memories, yet their short-term memory is preserved (*8*). But how are memories stored? It is hypothesized that information is stored in the brain as changes in strength of the connections (synapses) between neurons (*9*, *10*). Such changes in synaptic strength are observed when neuronal activity is recorded in the brain with microelectrodes: Relatively weak or infrequent stimulation elicits a short-lasting effect [early long-term potentiation (E-LTP)], whereas stronger or repeated stimulation elicits a sustained effect [late long-term potentiation (L-LTP)], lasting many hours instead of minutes. Similar to long-term memories, long-lasting changes in synaptic strength (L-LTP) are prevented by blocking protein synthesis (*2*, *5*–*7*, *11*, *12*).

If making new proteins is the rate-limiting step required to store new long-lasting memories, how is this process turned on? If one were able to identify the triggering mechanism and switch it on, then stimulation normally eliciting short-lasting changes should evoke long-lasting ones. Could an increase in the ability to make new proteins explain extraordinarily long-lasting memories?

Our work on the translation initiation factor eIF2α suggests that this could be the case (*13*): eIF2α regulates two fundamental processes that are essential for the generation of long-lasting memories. Chemical modification (phosphorylation) enhances the activity of eIF2α and suppresses general protein synthesis (*14*, *15*). Paradoxically, it also selectively stimulates the synthesis of a specific protein, ATF4 (CREB2) (*16*), a well-known memory repressor that blocks the new expression of genes needed for memories (*2*, *17*, *18*). In agreement with our idea, the activity of eIF2α is reduced in neurons exposed to L-LTP–inducing repeated stimulation, both increasing new protein synthesis and decreasing levels of the repressor (*13*, *19*).

We predicted that if we reduced eIF2α activity (phosphorylation), animals would be able to generate long-term memories from stimuli that normally generate only a short-term change. Thus, we studied mice in which the activity of eIF2α is reduced genetically (eIF2α$^{+/S51A}$ mice) (*13*). Orienting oneself in a particular space is a function of the hippocampus, a brain region that is crucial especially for spatial learning and memory. The spatial memory of the eIF2α$^{+/S51A}$ mice was assessed in the Morris water maze (*20*). In this task, mice swimming in a pool of opaque water search for a submerged platform. To this end, they use visual cues that are placed on the wall of the testing room to remember the location of the hidden platform. Remarkably, we found that after initial training in the pool, eIF2α$^{+/S51A}$ mice found the platform faster, indicating that their spatial memory was enhanced (see the figure, panel A). Moreover, the eIF2α$^{+/S51A}$

Eppendorf and *Science* are pleased to present the prize-winning essay by Mauro Costa-Mattioli, the 2008 winner of the Eppendorf and *Science* Prize for Neurobiology.



Control of eIF2α activity is critical for making long-term memories. (**A**) eIF2α$^{+/S51A}$ mice exhibit enhanced long-term spatial memory in the Morris water maze. (**B**) A weak synaptic stimulation elicits a long-lasting response in hippocampal slices from eIF2α$^{+/S51A}$ mice. (**C**) Injection of Sal003 into the hippocampus immediately after training blocks long-term spatial memory. Dark syringes refer to either vehicle or Sal003 infusions across groups. (**D**) The induction of L-LTP is blocked by Sal003. Latency is the time required for the mice to find the submerged platform. fEPSP (field excitatory postsynaptic potential) measures changes in synaptic strength. [Figure reprinted from (*13*) with permission from Elsevier]

Department of Neuroscience, Baylor College of Medicine, Houston, Texas, 77030. E-mail: costamat@bcm.edu

mice exhibited enhanced learning and memory in other behavioral tests, such as the ability to learn that an auditory or visual stimulus predicts a foot shock and that a specific taste precedes nausea (*13*).

We next asked whether these mice also had a greater ability to generate long-lasting changes in synaptic strength. Indeed, weak synaptic stimulation in hippocampal slices from $eIF2\alpha^{+/S51A}$ mice, which usually induces short-lasting changes in synaptic strength, generated L-LTP (see the figure, panel B). Accordingly, in another study we found that mice lacking the enzyme that activates eIF2α, GCN2, exhibit a similar phenotype (*19*). These data strongly support the idea that a reduction in the activity of eIF2α enhances LTP and mnemonic processes.

Would an increase in the activity of eIF2α have an opposite effect, that is, prevent long-lasting synaptic changes and memory storage? We tested this idea with Sal003, a drug that stimulates the activity of eIF2α by blocking eIF2α phosphatases (*13*). The Sal003-induced increase in eIF2α activity not only resulted in the inhibition of general protein synthesis but also increased the expression of ATF4 protein. Remarkably, animals injected with Sal003 in the hippocampus were unable to find the platform in the Morris water maze (see the figure, panel C) and they lost the ability to freeze in response to fear (*13*), clear indications that their memory was impaired. In addition, Sal003 blocked the long-lasting changes elicited by repeated synaptic activation (see the figure, panel D). The impaired ability to induce a sustained LTP was due to an increase in ATF4 protein levels, since L-LTP evoked in slices from mice lacking ATF4 was resistant to Sal003.

The ability to enhance memory formation by increasing new protein synthesis appears to be a widely conserved mechanism from sea slugs to mammals. Decreasing the activity of ATF4 in mice or ApCREB2 (the ortholog of ATF4) in the sea slug *Aplysia* lowers the threshold for long-lasting changes and memory (*17*, *18*).

By combining cellular, molecular, neurophysiological, and behavioral methods, our experiments reveal a crucial step in memory processing: The activity of eIF2α "decides" whether a long-term memory is made from an experience. It therefore remains essential to identify the genes regulated by eIF2α and determine if de novo mutations in these genes underlie eidetic and other enhanced forms of memory.

As "Ireneo Funes died in 1889" (*21*), we shall never know whether eIF2α activity was exceptionally low in his brain; however, it remains a possibility.

2008 Grand Prize Winner

The author of the prize-winning essay, **Mauro Costa-Mattioli**, received his bachelor's degree in biology from the University of the Republic, Montevideo, Uruguay. In 1998, he moved to France, where he received his master's degree (diplôme universitaire) from Pierre and Marie Curie University, Paris, and his Ph.D. from the University of Nantes under the supervision of Sylviane Billaudel. In 2002, he joined the laboratory of Nahum Sonenberg at McGill University, Montreal, as a postdoctoral fellow. His work defined the role of translational (protein synthesis) control in long-lasting synaptic plasticity and memory formation. In the summer of 2008, he joined the faculty at Baylor College of Medicine in Houston, Texas, as an assistant professor of neuroscience. Using multidisciplinary approaches, Dr. Costa-Mattioli's laboratory studies the molecular and cellular mechanisms underlying long-term synaptic plasticity, learning and memory, and related neurological disorders.

Finalists

Hendrikje Nienborg for her essay "Visual Perception: Interactions between Sensory and Decision Processes." Dr. Nienborg received an M.Sc. in neuroscience from the University of Oxford, UK, in 2000. She continued her studies at Munich University, Germany, and was awarded her M.D. in 2003 and her Ph.D. in 2005. She then joined Bruce Cumming's lab at the NIH as a postdoctoral fellow. Also a sculptor, Dr. Nienborg is fascinated by visual perception and our ability to perceive depth, the focus of her research. For her graduate work she showed that disparity-selective neurons in primate primary visual cortex are limiting factors for depth perception. In her postdoctoral work, she seeks to identify the mechanisms underlying neuronal correlates of perceptual decisions in behaving monkeys.

Claudio Hetz for his essay "Protein Misfolding Disorders and ER Stress Signals." Dr. Hetz received his undergraduate degree in biotechnology engineering from the University of Chile in 2000. In his Ph.D. work with Claudio Soto at Serono Pharmaceutical Research Institute, Geneva, he showed that prion pathogenesis involves endoplasmic reticulum stress responses. In post-doctoral work, first in Stanley Korsmeyer's laboratory at Dana-Farber Cancer Institute and then in Laurie Glimcher's lab at Harvard, he addressed the connection between apoptosis and the unfolded protein response in vivo. Since 2007 he has been an assistant professor at the University of Chile and an adjunct professor at Harvard. His laboratory uses animal models to investigate the signaling responses involved in protein misfolding disorders.

For the full text of finalist essays and for information about applying for next year's awards, see *Science* Online at www.sciencemag.org/feature/data/prizes/eppendorf/eppenprize.shtml.

CREDIT: (TOP) FEDERICO CIVINARI

References and Notes

1. E. S. Parker, L. Cahill, J. L. McGaugh, *Neurocase* **12**, 35 (2006).
2. E. R. Kandel, *Science* **294**, 1030 (2001).
3. J. L. McGaugh, *Science* **287**, 248 (2000).
4. R. L. Davis, *Annu. Rev. Neurosci.* **28**, 275 (2005).
5. R. J. Kelleher 3rd, A. Govindarajan, S. Tonegawa, *Neuron* **44**, 59 (2004).
6. E. Klann, T. E. Dever, *Nat. Rev. Neurosci.* **5**, 931 (2004).
7. M. A. Sutton, E. M. Schuman, *Cell* **127**, 49 (2006).
8. H. P. Davis, L. R. Squire, *Psychol. Bull.* **96**, 518 (1984).
9. T. V. Bliss, G. L. Collingridge, *Nature* **361**, 31 (1993).
10. R. C. Malenka, M. F. Bear, *Neuron* **44**, 5 (2004).
11. M. Costa-Mattioli, N. Sonenberg, *Prog. Brain Res.* **169**, 81 (2008).
12. O. Steward, E. M. Schuman, *Annu. Rev. Neurosci.* **24**, 299 (2001).
13. M. Costa-Mattioli *et al.*, *Cell* **129**, 195 (2007).
14. T. E. Dever, A. C. Dar, F. Sicheri, in *Translational Control in Biology and Medicine*, M. B. Mathews, N. Sonenberg, J. W. B. Hershey, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007), pp. 319–345.
15. T. V. Pestova, J. R. Lorsch, C. U. T. Hellen, in *Translational Control in Biology and Medicine*, M. B. Mathews, N. Sonenberg, J. W. B. Hershey, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007), pp. 87–128.
16. D. Ron, H. P. Harding, in *Translational Control in Biology and Medicine*, M. B. Mathews, N. Sonenberg, J. W. B. Hershey, Eds. (Cold Spring Harbor Laboratory Press, "Cold Spring Harbor, NY, 2007), pp. 345–368.
17. D. Bartsch *et al.*, *Cell* **83**, 979 (1995).
18. A. Chen *et al.*, *Neuron* **39**, 655 (2003).
19. M. Costa-Mattioli *et al.*, *Nature* **436**, 1166 (2005).
20. R. G. Morris, P. Garrud, J. N. Rawlins, J. O'Keefe, *Nature* **297**, 681 (1982).
21. J. L. Borges, in *Ficciones*, A. Kerrigan, Ed. (Grove, New York, 1962), pp. 107–116.

22. I thank my collaborators for their contribution to this work.

10.1126/science.1167228

Consciousness and Anesthesia

Michael T. Alkire,[1] Anthony G. Hudetz,[2] Giulio Tononi[3]*

When we are anesthetized, we expect consciousness to vanish. But does it always? Although anesthesia undoubtedly induces unresponsiveness and amnesia, the extent to which it causes unconsciousness is harder to establish. For instance, certain anesthetics act on areas of the brain's cortex near the midline and abolish behavioral responsiveness, but not necessarily consciousness. Unconsciousness is likely to ensue when a complex of brain regions in the posterior parietal area is inactivated. Consciousness vanishes when anesthetics produce functional disconnection in this posterior complex, interrupting cortical communication and causing a loss of integration; or when they lead to bistable, stereotypic responses, causing a loss of information capacity. Thus, anesthetics seem to cause unconsciousness when they block the brain's ability to integrate information.

How consciousness arises in the brain remains unknown. Yet, for nearly two centuries our ignorance has not hampered the use of general anesthesia for routinely extinguishing consciousness during surgery. Unfortunately, once in every 1000 to 2000 operations a patient may temporarily regain consciousness or even remain conscious during surgery (*1*). Such intraoperative awareness arises in part because our ability to evaluate levels of consciousness remains limited. Nevertheless, progress is being made in identifying general principles that underlie how anesthetics bring about unconsciousness (*2*–*6*) and how, occasionally, they may fail to do so.

Cellular Actions of Anesthetics

The cellular and molecular pharmacology of anesthetics has been reviewed extensively (*6*–*8*). General anesthetics fall into two main classes: intravenous agents used to induce anesthesia, generally administered together with sedatives or narcotics; and volatile agents, generally used for anesthesia maintenance (Table 1). Anesthetics are thought to work by interacting with ion channels that regulate synaptic transmission and membrane potentials in key regions of the brain and spinal cord. These ion-channel targets are differentially sensitive to various anesthetic agents (Table 1).

Anesthetics hyperpolarize neurons by increasing inhibition or decreasing excitation (*9*) and alter neuronal activity: The sustained firing typical of the aroused brain changes to a bistable burst-pause pattern (*10*) that is also observed in non–rapid-eye-movement (NREM) sleep. At intermediate anesthetic concentrations, neurons begin oscillating, roughly once a second, between a depolarized up-state and a hyperpolarized down-state (*11*). The up-state is similar to the sustained depolarization of wakefulness. The down-state shows complete cessation of synaptic activity for a tenth of a second or more, after which neurons revert to another up-state. As anesthetic doses increase, the up-state turns to a short burst and the down-state becomes progressively longer. These changes in neuronal firing patterns are reflected in the electroencephalogram (EEG) (electrical recording from the scalp) as a transition from the low-voltage, high-frequency pattern of wakefulness (known as activated EEG), to the slow-wave EEG of deep NREM sleep, and finally to an EEG burst-suppression pattern (*12*).

The Anesthetized Patient: Unconscious or Unresponsive?

Clinically, at low-sedative doses anesthetics cause a state similar to drunkenness, with analgesia, amnesia, distorted time perception, depersonalization, and increased sleepiness. At slightly higher doses, a patient fails to move in response to a command and is considered unconscious. This behavioral definition of unconsciousness, which was introduced with anesthesia over 160 years ago, while convenient, has drawbacks. For instance, unresponsiveness can occur without unconsciousness. When we dream, we have vivid conscious experiences, but are unresponsive because inhibition by the brainstem induces muscle paralysis (*13*). Similarly, paralyzing agents used to prevent unwanted movements during anesthesia do not remove consciousness (*14*).

Certain anesthetics may impair a person's willfulness to respond by affecting brain regions where executive decisions are made. This is not an issue for anesthetics that globally deactivate the brain, but it may be problematic for dissociative anesthetics like ketamine. Low doses of ketamine cause depersonalization, out-of-body experiences, forgetfulness, and loss of motivation to follow commands (*15*). At higher doses, ketamine causes a characteristic state in which the eyes are open and the face takes on a disconnected blank stare. Neuroimaging data show a complex pattern of regional metabolic changes (*16*), including a deactivation of executive circuits in anterior cingulate cortex and basal ganglia (Fig. 1) (*17*). A similar open-eyed unresponsiveness is seen in akinetic mutism after bilateral lesions around the anterior cingulate cortex (*18*). In at least some of these cases, patients understand questions, but may fail to respond. Indeed, a woman with large frontal lesions who was clinically unresponsive was asked to imagine playing tennis or to navigate her room, and she showed cortical activation patterns indistinguishable from those of healthy subjects (*19*). Thus, clinical unresponsiveness is not necessarily synonymous with unconsciousness.

At doses near the unconsciousness threshold, some anesthetics block working memory (*20*). Thus, patients may fail to respond because they immediately forget what to do. At much lower doses, anesthetics cause profound amnesia. Studies with the isolated forearm technique, in which a tourniquet is applied to the arm before paralysis is

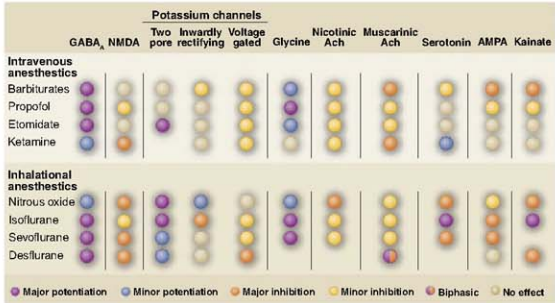


Table 1. Ionic mechanisms and targets of current clinical anesthetics (*6*, *8*). Abbreviations: Ach, acetylcholine; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; GABA$_A$, γ-aminobutyric acid, type A; NMDA, *N*-methyl-D-aspartate.

[1]Department of Anesthesiology and the Center for the Neurobiology of Learning and Memory, University of California, Irvine, CA 92868, USA. [2]Department of Anesthesiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [3]Department of Psychiatry, University of Wisconsin, Madison, WI 53719, USA.

*To whom correspondence should be addressed. E-mail: gtononi@wisc.edu

induced (to allow the hand to move while the rest of the body is paralyzed), show that patients under general anesthesia can sometimes carry on a conversation using hand signals, but postoperatively deny ever being awake (*21*). Thus, retrospective oblivion is no proof of unconsciousness.

Nevertheless, at some level of anesthesia between behavioral unresponsiveness and the induction of a flat EEG [indicating the cessation of the brain's electrical activity, one of the criteria for brain death (*22*)], consciousness must vanish. Therefore, the use of brain-function monitors could improve consciousness assessment during anesthesia (*23*). For instance, bispectral index monitors record the EEG signal over the forehead and reduce the complex signal into a single number that tracks a patient's depth of anesthesia over time (*12*). Such devices help guide anesthetic delivery and may reduce cases of intraoperative awareness (*24*), but they remain limited at directly indicating the presence or absence of consciousness, especially around the transition point. The isolated forearm technique has shown that individual patients can be aware and responsive during surgery even though their bispectral index value suggests they are not (*25*). Either the EEG is not sensitive enough to the neural processes underlying consciousness, or we still do not yet fully understand what to look for.

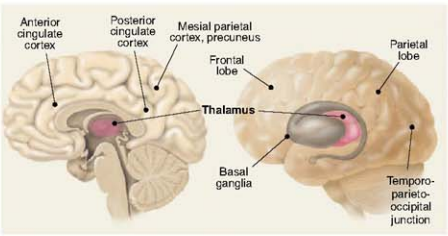


Fig. 1. Brain areas associated with anesthetic effects [see text and (*2*)].

The Thalamus—Switch or Readout?

The most consistent regional effect produced by anesthetics at (or near) loss of consciousness is a reduction of thalamic metabolism and blood flow (Fig. 1), suggesting that the thalamus may serve as a consciousness switch (*2*). Indeed, switchlike effects have been found with a number of thalamic manipulations. For example, γ-aminobutyric acid (GABA) agonists (mimicking anesthetic action) injected into the intralaminar nuclei cause rats to rapidly fall asleep, with a corresponding slowing of the EEG (*26*). Conversely, rats under anesthetic concentrations of sevoflurane can be awakened by a minute injection of nicotine into the intralaminar thalamus (*27*). In humans, midline thalamic damage can result in a vegetative state (*18*). Conversely, recovery from the vegetative state is heralded by the restoration of functional connectivity between thalamus and cingulate cortex (*28*). Also, deep brain electrical stimulation of the central thalamus improved behavioral responsiveness in a patient who was minimally conscious (*29*).

Nevertheless, thalamic activity does not decrease with all anesthetics. Ketamine increases global metabolism, especially in the thalamus (*16*). Other anesthetics can substantially reduce thalamic activity at doses that cause sedation, not unconsciousness. For instance, sevoflurane sedation causes a 23% reduction of relative thalamic metabolism when subjects are still awake and responsive (*30*). Indeed, anesthetic effects on the thalamus may be largely indirect (*6*, *31*, *32*). Spontaneous thalamic firing during anesthesia is largely driven by feedback from cortical neurons (*33*), especially anesthetic-sensitive layer V cells (*34*). Many of these cells also project onto brainstem arousal centers, so cortical deactivation can reduce both thalamic activity and arousal (*35*). Also, the metabolic and electrophysiological effects of anesthetics on the thalamus in animals are abolished by removal of the cortex (*33*, *34*, *36*). By contrast, after thalamic ablation, the cortex still produces an activated EEG (*37*), suggesting that the thalamus is not the sole mediator of cortical arousal, nor perhaps is it the most direct one. In patients with implanted brain electrodes undergoing a second surgery to place a deep brain stimulator, the cortical EEG changed dramatically the instant the patients lost consciousness (*38*). However, there was little change in thalamic EEG activity until 10 min later. Conversely, in epileptic patients, during REM sleep (usually associated with dreaming) the cortical EEG was activated as if patients were awake, but the thalamic EEG showed slow wave activity, as if patients were asleep (*39*). Thus, the effects of anesthetics on the thalamus may represent a readout of global cortical activity rather than a consciousness switch, and thalamic activity may not be a sufficient basis for consciousness.

Nonetheless, it is premature to write off the thalamus altogether. Perhaps efficient communication among cortical areas requires a thalamic relay (*40*), in which case thalamic lesions would lead to a functional disconnection despite an activated cortex. A functional thalamic disconnection during anesthesia has been found with neuroimaging (*41*). Subthreshold depolarization to many cortical areas may be provided by calbindin-positive matrix cells, which are especially concentrated within some intralaminar thalamic nuclei and project diffusely to superficial layers of cortex (*42*). Cells in intralaminar nuclei can fire at high frequencies, thus providing a coherent oscillatory bias that may facilitate long-range cortico-cortical interactions. Therefore, whereas cortical arousal may occur without the thalamus, consciousness may not.

Cortical Effects of Anesthetics

Are some cortical areas more important than others for the induction of unconsciousness by anesthetics? Evoked responses in primary sensory cortices—the first relay for incoming stimuli—are often unchanged during anesthesia, deep sleep, and in vegetative patients. Also, activity in primary sensory areas often does not correlate with perceptual experience (*43*). Frontal cortex too may not be essential for anesthetic unconsciousness, because different anesthetics have variable effects on this area. For instance, at equivalent hypnotic doses, both propofol and thiopental deactivate posterior brain areas, but only propofol deactivates frontal cortex (*44*). Furthermore, large lesions of the frontal cortex do not by themselves produce unconsciousness (*45*).

Anesthetic-induced unconsciousness is usually associated with deactivation of mesial parietal cortex, posterior cingulate cortex, and precuneus (Fig. 1) (*46*). These same areas are deactivated in vegetative patients but are the first to reactivate in those who recover (*28*). Moreover, neural activity in these areas is altered during seizures associated with an impairment of consciousness (*47*) and in sleep (*48*). These mesial cortical areas are strategically located at the main hub of the brain's connectional core (*49*). They are also part of a default network that is especially active at rest and may be involved in global monitoring of the internal environment and several functions related to the self (*50*). Nevertheless, mesial cortical areas are deactivated in REM sleep (*48*), when subjects experience vivid dreams. Intriguingly, at intermediate doses, certain anesthetics, such as nitrous oxide, produce a fairly selective deactivation of posterior mesial cortex (*51*), yet when these areas start to turn off, subjects report dreamlike feelings with depersonalization and out-of-body experiences, rather than unconsciousness.

In addition to mesial cortical areas, many anesthetics also deactivate or disconnect a lateral temporo-parieto-occipital complex of multimodal associative areas centered on the inferior parietal cortex (Fig. 1). In this case, lesion and anesthesia data are mutually supportive: Patients with bilateral lesions at the temporo-parieto-occipital junction show no sign of perceptual experience, despite a flurry of undirected motor activity, a condition called hyperkinetic mutism (*18*). Thus, a complex of posterior brain areas comprising the lateral temporo-parieto-occipital junction and perhaps a mesial cortical core are most likely the final common target for anesthetic-induced unconsciousness.

Disruption of Cortical Integration

Loss of consciousness may not necessarily require that neurons in these posterior brain areas be inactivated. Instead, it may be sufficient that dynamic aspects of neural activity change, especially if these affect the brain's ability to integrate information (Fig. 2) (*3*, *5*).

Consider first large-scale integration, loosely defined as the ability of different cortical regions to interact effectively (*52*). When consciousness fades during anesthesia, there is a drop in EEG coherence in the γ-frequency range (20 to 80 Hz) between right and left frontal cortices and between frontal and occipital regions (*4*). Anesthetics also suppress fronto-occipital γ coherence in animals, both under visual stimulation and at rest (*53*). The effect is gradual and much stronger for long-range than for local coherence (*53*). Anesthetics may disrupt cortical integration (*5*) by acting on structures that facilitate long-range cortico-cortical interactions, such as the posterior cortical connectional hub (*49*), certain thalamic nuclei (*42*), or possibly the claustrum (*54*). Anesthetics may also disrupt synchronization among distant areas by slowing neural responses (*55*).

The loss of feedback interactions in the cortex may be especially critical. When rats become unresponsive under anesthesia, information transfer first decreases in the feedback direction (Fig. 2B) (*56*). Also, anesthesia suppresses the late component (>100 ms) of visual responses, possibly by inhibiting feedback connections (*57*), but not the early feedforward components. Moreover, anesthesia abolishes contextual and attentional modulation of firing, presumably mediated by feedback connections (*58*). The corticothalamic system may be especially vulnerable to anesthetics due to its small-world organization. Small-world networks have mostly local connectivity with comparatively few long-range connections. Augmented with hubs, such networks maximize interactions while minimizing wiring. By the same token, anesthetics need only disrupt a few long-range connections to produce a set of disconnected components. Indeed, computer simulations demonstrate a rapid state transition at a critical anesthetic dose (*59*), consistent with a breakdown in network integration.

Disruption of Cortical Information Capacity

Consider next how anesthetics affect information, defined loosely as the number of discriminable activity patterns. When the repertoire of discriminable firing patterns available to the corticothalamic system shrinks, neural activity becomes less informative, even though it may be globally integrated (*52*). As described above, at high enough doses several anesthetics produce a burst-suppression pattern in which a near-flat EEG is interrupted every few seconds by brief, quasi-periodic bursts of global activation—a stereotypic, global on-off pattern. Such stereotypic burst-suppression can also be elicited by visual, auditory, and mechanical stimuli (Fig. 3B) (*60*, *61*). Thus, during deep anesthetic unconsciousness, the corticothalamic system can still be active—in fact, hyperexcitable—and can produce global responses. However, the repertoire of responses has shrunk to a stereotypic burst-suppression pattern, with a corresponding loss of information, essentially creating a system having only two possible states (on or off). Generalized convulsive seizures provide another example in which consciousness can be lost even though neural activity remains high and highly

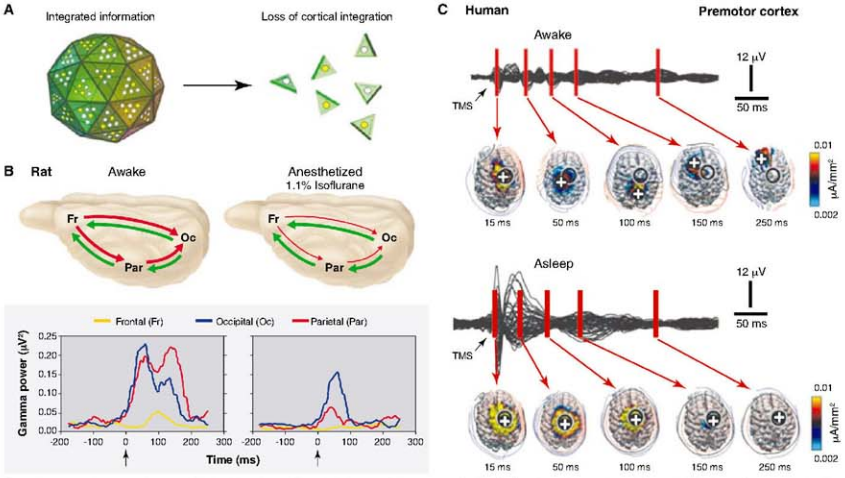


Fig. 2. Unconsciousness is associated with a loss of cortical integration. **(A)** The corticothalamic system is represented metaphorically as a large die having many faces, each corresponding to a different brain firing pattern. During conscious waking, the die rolls on a particular face, ruling out all the others and thus generating integrated information. If integration is lost (as in anesthesia or sleep), the die disintegrates into many two-faced dice, each generating 1 bit of information. **(B)** Anesthesia reduces cortical integration in the rat. (Top) During waking, transfer entropy, a measure of directional interactions among brain areas, is balanced in the feedforward (green) and feedback (red) directions. During anesthesia, feedback transfer entropy (red) is reduced, implying a decrease in front-to-back interactions. (Bottom) Responses to a flashing light delivered at 0.2 Hz (arrow) from a representative rat when awake and under 1.1% isoflurane anesthesia (*56*). When the rat is awake, each flash evokes a sustained γ-frequency (20 to 60 Hz) response in visual occipital cortex (blue) and a later response in parietal association cortex (red). During anesthesia, the occipital response is preserved, although it is shorter (blue), and the parietal response is attenuated, indicating that anesthesia reduces cortical interactions and thus reduces integration. **(C)** Sleeping reduces cortical integration in humans. EEG voltages and current densities are shown from a representative subject in which the premotor cortex was stimulated with transcranial magnetic stimulation (TMS) (black arrow). During waking (top), stimulation evokes EEG responses first near the stimulation site (black circle; the white cross is the site of maximum evoked current) and then in sequence at other cortical locations. During deep sleep (bottom), the stimulus-evoked response remains local, indicating a loss of cortical integration.

synchronized: A large portion of the corticothalamic complex is engaged in strong, hypersynchronous activity, but this activity is stereotypic (*60*, *61*).

A Bit Like Sleep

Sleep is the only time when healthy humans regularly lose consciousness. Subjects awakened during slow wave sleep early in the night may report short, thoughtlike fragments of experience, or often nothing at all (*13*). Although anesthesia is not the same as natural sleep, brain-arousal systems are similarly deactivated (*6*, *62*). Also, as under anesthesia, during slow wave sleep, cortical and thalamic neurons become bistable and undergo slow oscillations (1 Hz or less) between up- and down-states. Like animal studies during anesthesia (Fig. 2B and 3B), human studies during slow wave sleep suggest that the bistability of cortical neurons has consequences for the brain's capacity to integrate information (Figs. 2C and 3C). During wakefulness, transcranial magnetic stimulation (TMS) applied to premotor cortex and other cortical areas induces a sustained response (300 ms) involving the sequential activation of specific brain areas, the identity of which depends upon the precise site of stimulation (*63*, *64*). During early NREM sleep, possibly due to the induction of a local down-state, TMS pulses produce instead a short (<150 ms) local response (*64*), suggesting a loss of integration. Intriguingly, TMS pulses to mesial parietal regions, overlying the main hub in the cortical connectional core (*49*), trigger a stereotypic, high-amplitude slow wave closely resembling spontaneous slow waves (*63*). This stereotypic response, presumably due to the simultaneous activation of the cortical connectional core and to the induction of a global down-state, reflects a limited repertoire of activity patterns and thus a loss of information.

Consciousness and Integrated Information

The evidence from anesthesia and sleep states (Fig. 2 and 3) converges to suggest that loss of consciousness is associated with a breakdown of cortical connectivity and thus of integration, or with a collapse of the repertoire of cortical activity patterns and thus of information (Figs. 2 and 3). Why should this be the case? A recent theory suggests a principled reason: Information and integration may be the very essence of consciousness (*52*). Classically, information is the reduction of uncertainty among alternatives: When a coin falls on one of its two sides, it provides 1 bit of information, whereas a die falling on one of six faces provides ~2.6 bits. But then having any conscious experience, even one of pure darkness, must be extraordinarily informative, because we could have had countless other experiences instead (think of all the frames of every possible movie). Having any experience is like throwing a die with a trillion faces and identifying which number came up (Figs. 2A and 3A). On the other hand, every experience is an integrated whole that cannot be subdivided into independent components. For example, with an intact brain you cannot

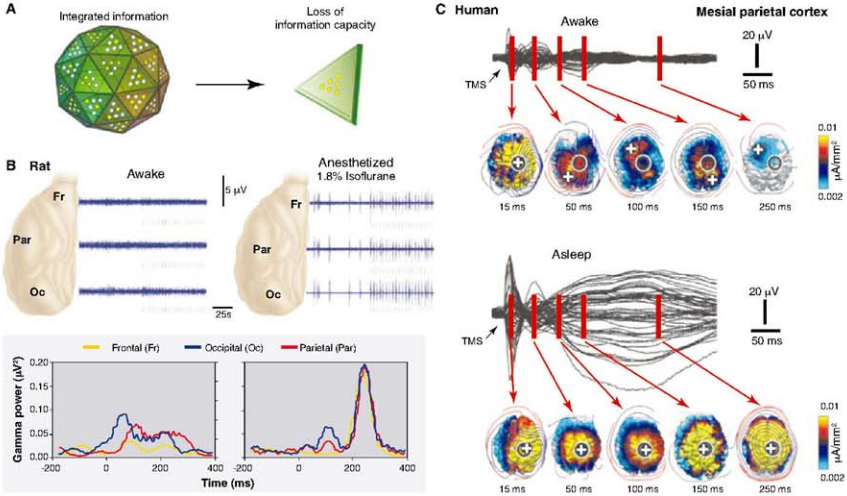


Fig. 3. Unconsciousness is associated with a loss of information capacity. **(A)** As in Fig 2, the corticothalamic system is represented metaphorically as a large die having many faces, each corresponding to a different brain firing pattern. During conscious waking, the die rolls on a particular face, ruling out all the others and thus generating integrated information. If information is lost (as in anesthesia or sleeping), the die is flattened so that it has only two faces (firing patterns). Due to the loss of repertoire, it generates only 1 bit of information. **(B)** Anesthesia reduces information capacity in rat cortex. (Top) Field potentials recorded before and during light flashes (marks below each trace). During waking (left), flash-evoked field potentials (blue) (light flashes indicated by marks below each trace) are small and variable, being masked by spontaneous neuronal activity. During deep anesthesia (right), bursts of activity occur spontaneously and after each light flash. (Bottom) During anesthesia, the γ-burst response is uniform across all three brain regions. Thus, responses are stereotypic and lack regional specificity, indicating a loss of information capacity. **(C)** Sleeping reduces cortical information carrying capacity in humans. (Top) During waking, stimulation over the mesial parietal cortex produces a specific, sequential pattern of activation. (Bottom) During sleep, stimulation produces a global, stereotypic response that spreads from the stimulation site to most of the cortex, indicating a loss of information capacity. Black traces represent averaged voltage potentials recorded at all electrodes and superimposed, while estimated current density is displayed in absolute scale (*63*, *64*).

experience the left half of the visual field independently of the right half, or visual shapes independently of their color. In other words, the die of experience is a single one—throwing multiple dice and combining the numbers will not do.

Less metaphorically, the theory claims that the level of consciousness of a physical system is related to the repertoire of different states (information) that can be discriminated by the system as a whole (integration). A measure of integrated information, called phi (Φ), can be used to quantify the information generated when a system enters one particular state of its repertoire, above and beyond the information generated independently by its parts (*52*, *65*). In practice, Φ can only be measured rigorously for small, simulated systems. However, empirical measures could be devised to evaluate integrated information on the basis of EEG data, resting functional connectivity, or TMS-evoked responses. This approach could allow the development of consciousness monitors that evaluate both loss of integration, as revealed by reduced functional or effective connectivity, and loss of information, as evidenced by stereotypic responses.

This theory has some interesting implications for anesthesia. For example, it explains why a corticothalamic complex is essential for consciousness and is thus the proper target for anesthesia: By conjoining functional specialization (each cortical area and neuronal group within each area is exquisitely specialized) with functional integration (thanks to extensive corticocortical and corticothalamocortical connectivity), a corticothalamic complex is well suited to behave as a single dynamic entity endowed with a large number of discriminable states. By contrast, parts of the brain made up of small, quasi-independent modules, such as the cerebellum, and parallel loops through the basal ganglia, are not sufficiently integrated, which is perhaps why they can be lesioned without loss of consciousness (*18*, *52*). The theory suggests that one should not interpret individual motor responses, or localized activations, as signs of consciousness, and conversely, should not interpret the absence of motor responses as a sure sign of unconsciousness. Finally, from this theoretical perspective, consciousness is not an all-or-none property, but it is graded: Specifically, it increases in proportion to a system's repertoire of discriminable states. The shrinking or dimming of the field of consciousness during sedation is consistent with this idea. On the other hand, the abrupt loss of consciousness at a critical concentration of anesthetics suggests that the integrated repertoire of neural states underlying consciousness may collapse nonlinearly.

Conclusions

Despite different mechanisms and sites of action, most anesthetic agents appear to cause unconsciousness by targeting, directly or indirectly, a posterior lateral corticothalamic complex centered around the inferior parietal lobe, and perhaps a medial cortical core. Whether the medial or lateral component is more important, and whether anterior cortical regions are critical primarily for executive functions and perhaps self-reflection, remain questions for future work. Second, anesthetics can cause unconsciousness not just by deactivating this posterior corticothalamic complex, but also by producing a functional disconnection between subregions of this complex. Third, although assessing loss of consciousness with verbal commands may usually be adequate, it may occasionally be misleading. Finally, one theoretical framework that seems to fit well with current empirical data suggests that consciousness requires an integrated system with a large repertoire of discriminable states. According to this framework, anesthetics would produce unconsciousness either by preventing integration (blocking the interactions among specialized brain regions) or by reducing information (shrinking the number of activity patterns available to cortical networks). Other frameworks for consciousness, emphasizing access to a global workspace (*66*, *67*), or the formation of large coalitions of neurons (*43*), are also consistent with many of the findings described here, especially those concerning the role of cortical integration. Together, these ideas should help in developing agents with more specific actions, in better monitoring their effects on consciousness, and in using anesthesia as a tool for characterizing the neural substrates of consciousness.

References and Notes

1. P. S. Sebel *et al.*, *Anesth. Analg.* **99**, 833 (2004).
2. M. T. Alkire, J. Miller, *Prog. Brain Res.* **150**, 229 (2005).
3. A. G. Hudetz, *Sem. Anesth. Perioper. Med. Pain* **25**, 196 (2006).
4. E. R. John, L. S. Prichep, *Anesthesiology* **102**, 447 (2005).
5. G. A. Mashour, *Anesthesiology* **100**, 428 (2004).
6. N. P. Franks, *Nat. Rev. Neurosci.* **9**, 370 (2008).
7. J. A. Campagna, K. W. Miller, S. A. Forman, *N. Engl. J. Med.* **348**, 2110 (2003).
8. U. Rudolph, B. Antkowiak, *Nat. Rev. Neurosci.* **5**, 709 (2004).
9. C. R. Ries, E. Puil, *J. Neurophysiol.* **81**, 1795 (1999).
10. R. R. Llinas, M. Steriade, *J. Neurophysiol.* **95**, 3297 (2006).
11. M. Steriade, I. Timofeev, F. Grenier, *J. Neurophysiol.* **85**, 1969 (2001).
12. L. Voss, J. Sleigh, *Best Pract. Res. Clin. Anaesthesiol.* **21**, 313 (2007).
13. J. A. Hobson, E. F. Pace-Schott, R. Stickgold, *Behav. Brain Sci.* **23**, 793 (2000).
14. G. P. Topulos, R. W. Lansing, R. B. Banzett, *J. Clin. Anesth.* **5**, 369 (1993).
15. M. R. Tucker, J. R. Hann, C. L. Phillips, *J. Oral Maxillofac. Surg.* **42**, 668 (1984).
16. J. W. Langsjo *et al.*, *Anesthesiology* **103**, 258 (2005).
17. M. T. Alkire, R. L. Gruver, J. W. Långsjö, K. Kaisti, H. Scheinin, *Anesthesiology* **104**, A1219 (2007).
18. J. B. Posner, C. B. Saper, N. D. Schiff, F. Plum, *Plum and Posner's Diagnosis of Stupor and Coma* (Oxford Univ. Press, Oxford, New York, ed. 4, 2007), Contemporary Neurology Ser. 71.
19. A. M. Owen *et al.*, *Science* **313**, 1402 (2006).
20. R. A. Veselis, R. A. Reinsel, V. A. Feshchenko, A. M. Dnistrian, *Anesthesiology* **97**, 329 (2002).
21. I. F. Russell, M. Wang, *Br. J. Anaesth.* **78**, 3 (1997).
22. E. F. Wijdicks, *N. Engl. J. Med.* **344**, 1215 (2001).
23. American Society of Anesthesiologists, *Anesthesiology* **104**, 847 (2006).
24. P. S. Myles, K. Leslie, J. McNeil, A. Forbes, M. T. Chan, *Lancet* **363**, 1757 (2004).
25. G. Schneider, A. W. Gelb, B. Schmeller, R. Tschakert, E. Kochs, *Br. J. Anaesth.* **91**, 329 (2003).
26. J. W. Miller, J. A. Ferrendelli, *Neuropharmacology* **29**, 649 (1990).
27. M. T. Alkire, J. R. McReynolds, E. L. Hahn, A. N. Trivedi, *Anesthesiology* **107**, 264 (2007).
28. S. Laureys, M. Boly, P. Maquet, *J. Clin. Invest.* **116**, 1823 (2006).
29. N. D. Schiff *et al.*, *Nature* **448**, 600 (2007).
30. M. T. Alkire *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1722 (2008).
31. M. T. Alkire, R. J. Haier, J. H. Fallon, *Conscious. Cogn.* **9**, 370 (2000).
32. N. D. Schiff, F. Plum, *J. Clin. Neurophysiol.* **17**, 438 (2000).
33. C. Vahle-Hinz, O. Detsch, M. Siemers, E. Kochs, *Exp. Brain Res.* **176**, 159 (2007).
34. A. Angel, *Br. J. Anaesth.* **71**, 148 (1993).
35. J. D. French, R. Hernandez-Peon, R. B. Livingston, *J. Neurophysiol.* **18**, 74 (1955).
36. K. Nakakimura, T. Sakabe, N. Funatsu, T. Maekawa, H. Takeshita, *Anesthesiology* **68**, 777 (1988).
37. J. Villablanca, M. E. Salinas-Zeballos, *Arch. Ital. Biol.* **110**, 383 (1972).
38. L. J. Velly *et al.*, *Anesthesiology* **107**, 202 (2007).
39. M. Magnin, H. Bastuji, L. Garcia-Larrea, F. Mauguiere, *Cereb. Cortex* **14**, 858 (2004).
40. R. W. Guillery, S. M. Sherman, *Neuron* **33**, 163 (2002).
41. N. S. White, M. T. Alkire, *Neuroimage* **19**, 402 (2003).
42. E. G. Jones, *Philos. Trans. R. Soc. London B Biol. Sci.* **357**, 1659 (2002).
43. F. Crick, C. Koch, *Nat. Neurosci.* **6**, 119 (2003).
44. R. A. Veselis *et al.*, *Anesth. Analg.* **99**, 399 (2004).
45. H. J. Markowitsch, J. Kessler, *Exp. Brain Res.* **133**, 94 (2000).
46. K. K. Kaisti *et al.*, *Anesthesiology* **96**, 1358 (2002).
47. H. Blumenfeld, *Prog. Brain Res.* **150**, 271 (2005).
48. P. Maquet, *J. Sleep Res.* **9**, 207 (2000).
49. P. Hagmann *et al.*, *PLoS Biol.* **6**, e159 (2008).
50. K. Vogeley *et al.*, *J. Cogn. Neurosci.* **16**, 817 (2004).
51. F. E. Gyulai, L. L. Firestone, M. A. Mintun, P. M. Winter, *Anesth. Analg.* **83**, 291 (1996).
52. G. Tononi, *BMC Neurosci.* **5**, 42 (2004).
53. O. A. Imas, K. M. Ropella, J. D. Wood, A. G. Hudetz, *Neurosci. Lett.* **402**, 216 (2006).
54. F. C. Crick, C. Koch, *Philos. Trans. R. Soc. London B Biol. Sci.* **360**, 1271 (2005).
55. R. Munglani, J. Andrade, D. J. Sapsford, A. Baddeley, J. G. Jones, *Br. J. Anaesth.* **71**, 633 (1993).
56. O. A. Imas, K. M. Ropella, B. D. Ward, J. D. Wood, A. G. Hudetz, *Neurosci. Lett.* **387**, 145 (2005).
57. A. G. Hudetz, *Int. Anesthesiol. Clin.* **46**, 25 (2008).
58. H. Super, H. Spekreijse, V. A. Lamme, *Nat. Neurosci.* **4**, 304 (2001).
59. M. L. Steyn-Ross, D. A. Steyn-Ross, J. W. Sleigh, L. C. Wilcocks, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 011917 (2001).
60. A. G. Hudetz, O. A. Imas, *Anesthesiology* **107**, 983 (2007).
61. D. Kroeger, F. Amzica, *J. Neurosci.* **27**, 10597 (2007).
62. R. Lydic, H. A. Baghdoyan, *Anesthesiology* **103**, 1268 (2005).
63. M. Massimini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8496 (2007).
64. M. Massimini *et al.*, *Science* **309**, 2228 (2005).
65. D. Balduzzi, G. Tononi, *PLoS Comput. Biol.* **4**, e1000091 (2008).
66. B. J. Baars, *Prog. Brain Res.* **150**, 45 (2005).
67. S. Dehaene, C. Sergent, J. P. Changeux, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8520 (2003).

68. Supported by the NIH Director's Pioneer Award and the James S. McDonnell Foundation (G.T.). G.T. has a patent pending on the use of TMS-EEG in anesthesia.

10.1126/science.1149213

Genetic Mapping in Human Disease

David Altshuler,[1,2,3,4,5]* Mark J. Daly,[1,2,5]* Eric S. Lander[1,6,7,8]*

Genetic mapping provides a powerful approach to identify genes and biological processes underlying any trait influenced by inheritance, including human diseases. We discuss the intellectual foundations of genetic mapping of Mendelian and complex traits in humans, examine lessons emerging from linkage analysis of Mendelian diseases and genome-wide association studies of common diseases, and discuss questions and challenges that lie ahead.

By the early 1900s, geneticists understood that Mendel's laws of inheritance underlie the transmission of genes in diploid organisms. They noted that some traits are inherited according to Mendel's ratios, as a result of alterations in single genes, and they developed methods to map the genes responsible. They also recognized that most naturally occurring trait variation, while showing strong correlation among relatives, involves the action of multiple genes and nongenetic factors.

Although it was clear that these insights applied to humans as much as to fruit flies, it took most of the century to turn these concepts into practical tools for discovering genes contributing to human diseases. Starting in the 1980s, the use of naturally occurring DNA variation as markers to trace inheritance in families led to the discovery of thousands of genes for rare Mendelian diseases. Despite great hopes, the approach proved unsuccessful for common forms of human diseases—such as diabetes, heart disease, and cancer—that show complex inheritance in the general population.

Over the past year, a new approach to genetic mapping has yielded the first general progress toward mapping loci that influence susceptibility to common human diseases. Still, most of the genes and mutations underlying these findings remain to be defined, let alone understood, and it remains unclear how much of the heritability of common disease they explain. Below, we discuss the intellectual foundations of genetic mapping, examine emerging lessons, and discuss questions and challenges that lie ahead.

Genetic Mapping by Linkage and Association

Genetic mapping is the localization of genes underlying phenotypes on the basis of correlation with DNA variation, without the need for prior hypotheses about biological function. The simplest form, called linkage analysis, was conceived by Sturtevant for fruit flies in 1913 (*1*). Linkage analysis involves crosses between parents that vary at a Mendelian trait and at many polymorphic variants ("markers"); because of meiotic recombination, any marker showing correlated segregation ("linkage") with the trait must lie nearby in the genome.

In the 1970s, the ability to clone and sequence DNA made it possible to tie genetic linkage maps in model organisms to the underlying DNA sequence, and thereby to molecularly clone the genes responsible for any Mendelian trait solely on the basis of their genomic position (*2*, *3*). Such studies typically involved three steps: (i) identifying the locus responsible through a genome-wide search; (ii) sequencing the region in cases and controls to define causal mutation(s); and (iii) studying the molecular and cellular functions of the genes discovered. So-called "positional cloning" became a mainstay of experimental genetics, identifying pathways that are crucial in development and physiology.

Linkage analysis in humans. For most of the 20th century, genome-wide linkage mapping was impractical in humans: Family sizes are small, crosses are not by design, and there were too few classical genetic markers to systematically trace inheritance. Progress in identifying the genes contributing to human traits was initially limited to studies of biological candidates such as blood-type antigens (*4*) and hemoglobin β protein in sickle-cell anemia (*5*).

In 1980, Botstein and colleagues, building on their use of DNA polymorphisms to study linkage in yeast (*6*) and the finding of DNA polymorphism at the globin locus in humans (*7*, *8*), proposed the use of naturally occurring DNA sequence polymorphisms as generic markers to create a human genetic map and systematically trace the transmission of chromosomal regions in families (*9*). The feasibility of genetic mapping in humans was soon demonstrated with the localization of Huntington disease in 1983 (*10*). A rudimentary genetic linkage map with ~400 DNA markers was generated by 1987 (*11*) and was fleshed out to ~5000 markers by 1996 (*12*). Physical maps providing access to linked chromosomal regions were developed by 1995 (*13*). With these tools, positional cloning became possible in humans, and the number of disorders tied to a specific gene grew from ~100 in the late 1980s to >2200 today (*14*).

Several lessons emerged from studies of Mendelian disease genes: (i) The "candidate gene" approach was woefully inadequate; most disease genes were completely unsuspected on the basis of previous knowledge. (ii) Disease-causing mutations often cause major changes in encoded proteins. (iii) Loci typically harbor many disease-causing alleles, mostly rare in the population. (iv) Mendelian diseases often revealed great complexity, such as locus heterogeneity, incomplete penetrance, and variable expressivity.

Geneticists were eager to apply genetic mapping to common diseases, which also show familial clustering. Mendelian subtypes of common diseases [such as breast cancer (*15*), hypertension (*16*), and diabetes (*17*)] were elucidated, but mutations in these genes explained few cases in the population. In common forms of common disease, risk to relatives is lower than in Mendelian cases, and linkage studies with excellent power to detect a single causal gene yielded equivocal results.

These features were consistent with, but did not prove, a polygenic model. The idea that commonly varying traits might be polygenic in nature was offered by East in 1910 (*18*). By 1920, linkage mapping was used to identify multiple unlinked factors influencing truncate wings in *Drosophila* (*19*), and Fisher had developed a mathematical framework for relating Mendelian factors and quantitative traits (*20*). In the late 1980s, linkage mapping of complex traits was made feasible for experimental organisms through the use of genetic mapping in large crosses (*21*). But there was little success in humans.

Genetic association in populations. A possible path forward emerged from population genetics and genomics. Instead of mapping disease genes by tracing transmission in families, one might localize them through association studies—that is, comparisons of frequencies of genetic variants among affected and unaffected individuals.

Genetic association studies were not a new idea. In the 1950s, such studies revealed correlations between blood-group antigens and peptic ulcer disease (*4*); in the 1960s and 1970s, common variation at the human leukocyte antigen (HLA) locus was associated with autoimmune and infectious diseases (*22*); and in the 1980s, apolipoprotein E was implicated in the etiology of Alzheimer's disease (*23*). Still, only about a dozen extensively reproduced associations of common variants (outside the HLA locus) were identified in the 20th century (*24*).

A central problem was that association studies of candidate genes were a shot in the dark: They were limited to specific variants in biological candidate genes, each with a tiny a priori probability of being disease-causing. Moreover, association studies were susceptible to false positives due to population structure, because there was no way to assess differences in the genetic background of cases and controls. Although many claims of associations were published, the statistical support tended to be weak and few were subsequently replicated (*25*).

In the mid-1990s, a systematic genome-wide approach to association studies was proposed (*26*–*28*): to develop a catalog of common human genetic variants and test the variants for associa-

[1]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [2]Center for Human Genetic Research and Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. [3]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA. [5]Department of Medicine, Harvard Medical School, Boston, MA 02114, USA. [6]Department of Systems Biology, Harvard Medical School, Boston, MA 02114, USA. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: altshuler@molbio.mgh.harvard.edu (D.A.); mjdaly@chgr.mgh.harvard.edu (M.J.D.); lander@broad.mit.edu (E.S.L.)

tion to disease risk. The focus on common variants as a mapping tool was a matter of practicality, grounded in population genetics. The human population has recently grown exponentially from a small size. As predicted by classical theory (*29*), humans have limited genetic variation: The heterozygosity rate for single-nucleotide polymorphisms (SNPs) is ~1 in 1000 bases (*30*–*32*). Moreover, perhaps 90% of heterozygous sites in each individual are common variants, typically shared among continental populations (*33*).

If most genetic variation in an individual is common, then why are mutations responsible for Mendelian diseases typically rare? One answer is natural selection: Mutations that cause strongly deleterious phenotypes—as most Mendelian diseases appear to be—are lost to purifying selection. But if deleterious mutations are typically rare, how could common variants play a role in disease? Common diseases often have late onset, with modest or no obvious impact on reproductive fitness. Mildly deleterious alleles can rise to moderate frequency, particularly in populations that have undergone recent expansion (*34*). Moreover, some alleles that were advantageous or neutral during human evolution might now confer susceptibility to disease because of changes in living conditions accompanying civilization. Finally, disease-causing alleles could be maintained at high frequency if they were under balancing selection, with disease burden offset by a beneficial phenotype (as in sickle-cell disease and malaria resistance).

These lines of reasoning led to the so-called "common disease–common variant" (CD-CV) hypothesis: the proposal that common polymorphisms (classically defined as having a minor allele frequency of >1%) might contribute to susceptibility to common diseases (*26*–*28*). If so, genome-wide association studies (GWASs) of common variants might be used to map loci contributing to common diseases. The concept

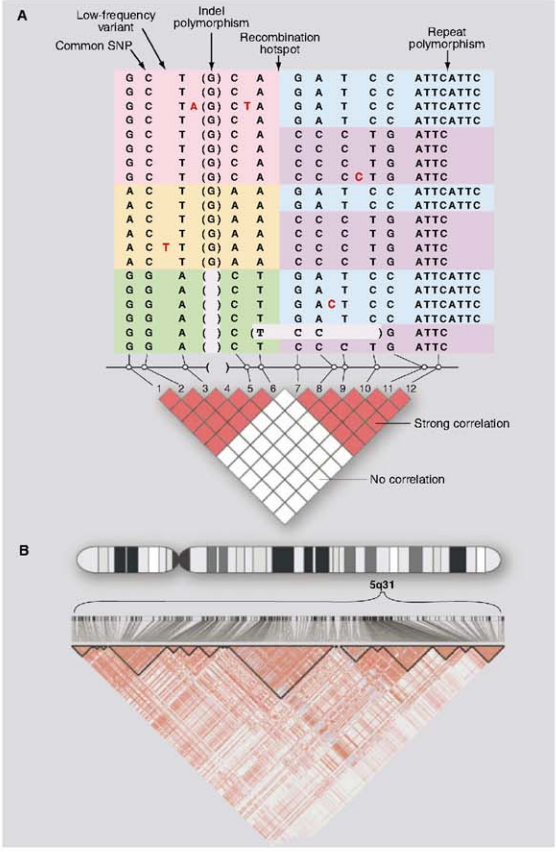


Fig. 1. DNA sequence variation in the human genome. **(A)** Common and rare genetic variation in 10 individuals, carrying 20 distinct copies of the human genome. The amount of variation shown here is typical for a 5-kb stretch of genome and is centered on a strong recombination hotspot. The 12 common variations include 10 SNPs, an insertion-deletion polymorphism (indel), and a tetranucleotide repeat polymorphism. The six common polymorphisms on the left side are strongly correlated. Although these six polymorphisms could theoretically occur in 2^6 possible patterns, only three patterns are observed (indicated by pink, orange, and green). These patterns are called haplotypes. Similarly, the six common polymorphisms on the right side are strongly correlated and reside on only two haplotypes (indicated by blue and purple). The haplotypes occur because there has not been much genetic recombination between the sites. By contrast, there is little correlation between the two groups of polymorphisms, because a hotspot of genetic recombination lies between them. The pairwise correlation between the common sites is shown by the red and white boxes below, with red indicating strong correlation and white indicating weak correlation. In addition to the common polymorphisms, lower-frequency polymorphisms also occur in the human genome. Five rare SNPs are shown, with the variant nucleotide marked in red and the reference nucleotide not shown. In addition, on the second to last chromosome, a larger deletion variant is observed that removes several kilobases of DNA. Such larger deletion or duplication events (i.e., CNVs) may be common and segregate as other DNA variants. **(B)** Small regions such as in (A) are often embedded in genomic regions with much greater extents of LD. The diagram shows actual data from the International HapMap Project, showing 420 genetic variants in a region of 500 kb on human chromosome 5q31. Positions of the variants and the pairwise correlations are shown below. Blocks of strong correlation are indicated by the black outlines. Longer-range patterns are often more complex than shown in (A) because weaker recombination hotspots may reduce, but not completely eliminate, marker-to-marker correlation.

was not that all causal mutations at these genes should be common (to the contrary, a full spectrum of alleles is expected), only that some common variants exist and could be used to pinpoint loci for detailed study.

It took a decade to develop the tools and methods required to test the CD-CV hypothesis: (i) catalogs of millions of common variants in the human population, (ii) techniques to genotype these variants in studies with thousands of patients, and (iii) an analytical framework to distinguish true associations from noise and artifacts.

Cataloging SNPs and linkage disequilibrium. Pilot projects in the late 1990s showed that it was possible to identify thousands of SNPs and to perform highly multiplexed genotyping by means of DNA microarrays (*35*). A public-private partnership, the SNP Consortium, built an initial map of 1.4 million SNPs (*32*); this has grown to more than 10 million SNPs (*36*) and is estimated to contain 80% of all SNPs with frequencies of >10% (*37*).

As the SNP catalog grew, a critical question loomed: Would GWASs require directly testing each of the ~10 million common variants for association to disease? That is, if only 5% of variants were tested, would 95% of associations be missed? Or could a subset serve as reliable proxies for their neighbors? Experience from Mendelian diseases suggested that substantial efficiencies might be possible. Each disease-causing mutation arises on a particular copy of the human genome and bears a specific set of common alleles in cis at nearby loci, termed a haplotype. Because the recombination rate is low [~1 crossover per 100 megabases (Mb) per generation], disease alleles in the population typically show association with nearby marker alleles for many generations, a phenomenon termed linkage disequilibrium (LD) (Fig. 1).

Early studies had demonstrated LD of nearby polymorphisms at the globin locus (*38*), which proved useful in tracking sickle-cell mutation. In the mid-1980s, it was proposed that a genome-wide search might be performed in genetically isolated populations, scanning the genome for a haplotype shared among unrelated patients carrying the same founder mutation (*39*). Such "LD mapping" in essence treated the entire population as a very large and very old extended family. This method soon proved useful in fine-mapping the founder Δ508 mutation in the transmembrane conductance regulator CFTR as a cause of cystic fibrosis (*40*) and in screening the entire genome in isolated populations such as Finland (*41*).

The key question was whether the same approach could be used more generally to study common alleles in large human populations, where recombination had more time to whittle down haplotypes. A simulation study suggested that LD might typically be too short to be useful, with a SNP every 5 kb (500,000 SNPs across the genome) providing very weak LD (average correlation $r^2 = 0.1$) (*42*). Studies of individual loci showed great heterogeneity in local LD (*43*).

As denser genetic maps became available, a clear picture emerged. Nearby variants were observed to form a block-like structure consisting of regions characterized by little evidence for historical recombination and limited haplotype diversity (*44, 45*). Within such regions, which soon proved general (*46*), genotypes of common SNPs could be inferred from knowledge of only a few empirically determined tag SNPs (*45–47*). These patterns were shaped by hot and cold spots of recombination in the human genome (*48–50*), as well as historical population bottlenecks (*51*).

The International HapMap Project was launched in 2002, with the goal of characterizing SNP frequencies and local LD patterns across the human genome in 270 samples from Europe, Asia, and West Africa. The project genotyped ~1 million SNPs by 2005 (*37*) and more than 3 million by 2007 (*52*). Sequence data collected by the project confirmed that the vast majority of common SNPs are strongly correlated to one or more nearby proxies: 500,000 SNPs provide excellent power to test >90% of common SNP variation in out-of-Africa populations, with roughly twice that number required in African populations (*37*).

Massively parallel genotyping. SNP genotyping was initially performed one SNP at a time, at a cost of ~$1 per measurement. Multiplex genotyping of hundreds of SNPs on DNA microarrays was demonstrated in 1998 (*35*), and capacity per array grew from 10,000 to 100,000 SNPs in 2002 to 500,000 to 1 million SNPs in 2007. In parallel, cost fell to $0.001 per genotype, or less than $1000 per sample for a whole-genome analysis. By 2006, several technologies could simultaneously genotype hundreds of thousands of SNPs at >99% completeness and >99% accuracy.

Copy-number variation. SNPs are only one type of genetic variation (Fig. 1). Using microarray technology, two groups in 2004 observed that individual copies of the human genome contain large regions (tens to hundreds of kilobases in size) that are deleted, duplicated, or inverted relative to the reference sequence (*53, 54*). Structural variants had been previously associated with developmental disorders and were often assumed to be pathogenic; the presence of so many segregating copy-number variations (CNVs) in the general population was surprising. The generality of CNVs was soon established (*55–59*). Many CNVs display tight LD with nearby SNPs (*56, 57*) and thus can be proxied by nearby SNPs in GWASs. Others occur in regions that are difficult to follow with SNPs, are highly mutable, or are rare (*58, 59*). Hybrid genotyping platforms have recently been developed to genotype SNPs and CNVs simultaneously (*60*).

Statistical analysis. Recognizing causal loci amid a genome's worth of random fluctuation required advances in statistical design, analysis, and interpretation. The risk of false negatives was illustrated by a study of type 2 diabetes (T2D) and the $Pro^{12} \rightarrow Ala$ polymorphism in peroxisome proliferator-activated receptor γ. Whereas an initial positive report (*61*) had not been confirmed in four modest-sized replication studies, larger studies produced strong and consistent evidence of increased risk by a factor of 1.2 (*62, 63*). The negative studies were actually consistent with the level of increased risk, but simply lacked adequate power to detect it.

Conversely, stringent thresholds for statistical significance are needed to avoid false positives due to multiple hypothesis testing. Simulations indicated that a dense genome-wide scan of common variants involves the equivalent of ~1 million independent hypotheses (*64*). A significance level of $P = 5 \times 10^{-8}$ thus represents a finding expected by chance once in 20 GWASs. Large sample sizes would be needed to reach such a stringent threshold (Fig. 2).

Systematic biases could also cause false positives. Differences in ancestry between cases and controls would yield spurious associations (*65*), suggesting the need for family-based controls (*66*). It was later recognized that genome-wide studies provide their own internal control: Mismatched ancestry is readily detectable because it produces frequency differences at thousands of SNPs, which could not all reflect causal associations. Methods were developed to detect and adjust for such biases (*67–69*) as well as unexpected relatedness between subjects. Technical artifacts, which are particularly problematic if cases and controls are not genotyped in parallel (*70*), were overcome by improved genotyping methods, quality control, and stringent filtering. To maximize efficiency and power, several groups developed methods of selecting tag SNPs (*47, 71–73*) from empirical LD data and using them to impute genotypes at other SNPs not genotyped in clinical samples (*74*) on the basis of LD relationships in the HapMap.

Genome-Wide Associations: Lessons

By early 2006, the tools were in place and studies were under way in many laboratories to resolve the hotly debated issue (*75, 76*) of whether genetic mapping of common SNPs would shed light on common disease. Since then, scores of publications have reported the localization of common SNPs associated with a wide range of common diseases and clinical conditions (age-related macular degeneration, type 1 and type 2 diabetes, obesity, inflammatory bowel disease, prostate cancer, breast cancer, colorectal cancer, rheumatoid arthritis, systemic lupus erythematosus, celiac disease, multiple sclerosis, atrial fibrillation, coronary disease, glaucoma, gallstones, asthma, and restless leg syndrome) as well as various individual traits (height, hair color, eye color, freckles, and HIV viral set point). Figure 3 illustrates data from a paradigmatic genome-wide association study of Crohn's disease performed by the Wellcome Trust Case Control Consortium.

Various lessons have already emerged about genetic mapping by GWAS:

1) GWASs work. Before 2006, only about two dozen reproducible associations outside the HLA locus had been discovered (*25*). By early 2008, more than 150 relationships were identified between common SNPs and disease traits (table S1). In most diseases studied, GWASs have revealed multiple independent loci, although some traits have not yet yielded associations that meet stringent thresholds (e.g., hypertension). It is not clear whether this

reflects inadequate sample size, phenotypic definition, or a different genetic architecture.

2) Effect sizes for common variants are typically modest. In a few cases, common variants with effects of a factor of ≥2 per allele have been found: APOE4 in Alzheimer's disease (*23*), CFH in age-related macular degeneration (*77–79*), and LOXL1 in exfoliative glaucoma (*80*). In the vast majority of cases, however, the estimated effects are much smaller—mostly increases in risk by a factor of 1.1 to 1.5 per associated allele.

3) The power to detect associations has been low. Given the effect sizes now known to exist, and the need to exceed stringent statistical thresholds, the first wave of GWASs provided low power to discover disease-causing loci (*81*, *82*). For example, achieving 90% power to detect an allele with 20% frequency and a factor of 1.2 effect at a statistical significance of 10^{-8} requires 8600 samples (Fig. 2). Thus, although it is unlikely that common alleles of large effect have been missed, GWASs of hundreds to several thousand cases have necessarily identified only a fraction of the loci that can be found with larger sample sizes. This prediction has been empirically confirmed in T2D (*83*), serum lipids (*84*, *85*), Crohn's disease (*86*), and height (*87–90*). Across these four traits and diseases, individual GWASs together documented 29 associations. Increasing the power by pooling the samples to perform meta-analysis and replication genotyping has increased this yield to more than 100 replicated loci for these four conditions.

4) Association signals have identified small regions for study but have not yet identified causal genes and mutations. Genetic mapping is a double-edged sword: Local correlation of genetic variants facilitates the initial identification of a region but makes it difficult to distinguish causal mutation(s). Luckily, whereas family-based linkage methods typically yield regions of 2 to 10 Mb in span, GWASs typically yield more manageable regions of 10 to 100 kb. These regions have yet to be scrutinized by fine-mapping and resequencing to identify the specific gene and variants responsible. Even when a locus is identified by SNP association, the causal mutation itself need not be a SNP. For example, the *IRGM* gene was associated with Crohn's disease on the basis of GWAS. Subsequent study suggests that the causal mutation is a deletion upstream of the promoter affecting tissue-specific expression (*91*).

5) A single locus can contain multiple independent common risk variants. Intensive study has already identified seven independent alleles at 8q24 for prostate cancer (*92*), three at complement factor H (CFH) for age-related macular degeneration (*93*, *94*), three at IRF5 for systemic lupus erythematosus (*95*), and two at IL23R for Crohn's disease (*96*). Multiple distinct alleles with different frequencies and risk ratios may well be the rule.

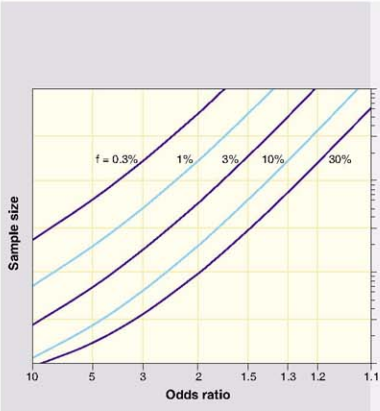
6) A single locus can harbor both common variants of weak effect and rare variants of large effect. In recent GWASs, studies of common SNPs enabled the identification of 19 loci as influencing low- or high-density lipoprotein (LDL, HDL) or triglycerides (*84*, *85*). Nine of these 19 were already known to carry rare Mendelian mutations with large effects, such as the loci for the LDL receptor (LDLR) and familial hypercholesterolemia (FH). Similarly, the genes encoding Kir6.2, WFS1, and TCF2 are all known to cause Mendelian syndromes including T2D, as well as common SNPs with modest effects.

7) Because allele frequencies vary across human populations, the relative roles of common susceptibility genes can vary among ethnic groups. One example is the association of prostate cancer at 8q24: SNPs in the region play a role in all ethnic groups, but the contribution is greater in African Americans. This is not because the risk alleles yet found confer greater susceptibility in African Americans, but because they occur at higher frequencies (*92*), contributing to the higher incidence among African American men than among men of European ancestry.

Lessons have also emerged about the functions and phenotypic associations of genes related to common diseases:

1) A subset of associations involve genes previously related to the disease. Of 19 loci meeting genome-wide significance in a recent GWAS of LDL, HDL, or triglyceride levels, 12 contained genes with known functions in lipid biology (*84*, *85*). The gene for 3-hydroxy-3-methyl glutaryl–coenzyme A reductase (HMGCR), encoding the rate-limiting enzyme in cholesterol biosynthesis and the target of statin medications, was found by GWAS to carry common genetic variation influencing LDL levels (*84*, *85*). Similarly, SNPs in the β-cell zinc transporter encoded by SLC30A8 were associated with risk of T2D (*97*).

2) Most associations do not involve previous candidate genes. In some cases, GWAS results immediately suggest new biological hypotheses—for example, the role of complement factor H in age-related macular degeneration (*77–79*), FGFR2 in breast cancer (*98*), and CDKN2A and CDKN2B in T2D (*99–101*). In many other cases, such as



GWAS			Rare variant	Nominal
90%	50%	10%	90%	90%
100,000	66,790	40,260	66,040	21,370
30,000	20,037	12,078	19,812	6411
10,000	6679	4026	6604	2137
3000	2004	1208	1981	641
1000	668	403	660	214
300	200	121	198	64
100	67	40	66	21

Fig. 2. Sample sizes required for genetic association studies. The graphs show the total number *N* of samples (consisting of *N*/2 cases and *N*/2 controls) required to map a genetic variant as a function of the increased risk due to the disease-causing allele (*x* axis) and the frequency of the disease-causing allele (various curves). The required sample size is shown in the table on the right for various different kinds of association studies. The first three columns pertain to GWASs using common variants across the entire genome; the columns correspond to different levels of statistical power to achieve a significant result at $P < 10^{-8}$. The fourth column pertains to a search for rare variants where the frequency listed is the collective frequency of rare variants in controls, and the odds ratio is the excess in cases as compared to controls. Sample sizes assume correction for a genome-wide search of ~20,000 protein-coding genes in the genome (aiming to achieve $P < 10^{-5}$ with one test performed per gene). The fifth column pertains to a test of a single hypothesis (e.g., testing association with a single SNP). For example, in a GWAS, 1000 samples provide 90% statistical power to detect a 30% allele with a factor of 2 effect. In a genome-wide search via exon sequencing, 660 samples provide 90% power to detect a gene in which rare variants have aggregate population frequency 1% and convey a factor of ~8 increase in risk. Note that the sample size to test essentially all common SNPs in the human genome is only 5 times the sample size to test a single SNP.

LOC387715/HTRA1 with age-related macular degeneration (*102*), nearby genes have no known function.

3) Many associations implicate non–protein-coding regions. Although some associated non-coding SNPs may ultimately prove attributable to LD with nearby coding mutations, many are sufficiently far from nearby exons to make this outcome unlikely. Examples include the region at 8q24 associated with prostate, breast, and colon cancer, 300 kb from the nearest gene (*103*, *104*), and the region at 9q21 associated with myocardial infarction and T2D, 150 kb from the nearest genes encoding CDKN2A and CDKN2B (*99*–*101*, *105*–*107*).

A role for noncoding sequence in disease risk is not surprising: Comparative genome analysis has shown that 5% of the human genome is evolutionarily conserved and thus functional; less than one-third of this 5% consists of genes that encode proteins (*108*). Noncoding mutations with roles in disease susceptibility will likely open new doors to understanding genome biology and gene regulation. Regulatory variation also suggests different therapeutic strategies: Modulating levels of gene expression may prove more tractable than replacing a fully defective protein or turning off a gain-of-function allele.

4) Some regions contain expected associations across diseases and traits. Crohn's disease, psoriasis, and ankylosing spondylitis have long been recognized to share clinical features; the association of the same common polymorphisms in IL23R in all three diseases points to a shared molecular cause (*96*, *109*, *110*). SNPs in STAT4 (signal transducer and activator of transcription 4) are associated with rheumatoid arthritis and systemic lupus, two diseases that share clinical features. Multiple variants associated with T2D are associated with insulin secretion defects in nondiabetic individuals (*101*, *111*–*116*), highlighting the role of β-cell failure in the pathogenesis of T2D.

5) Some regions reveal surprising associations. For example, unexpected connections have emerged among T2D, inflammatory diseases (two loci), and cancer (four loci). A single intron of CDKAL1 was found to contain a SNP associated with T2D and insulin secretion defects (*99*–*101*, *116*), and another with Crohn's disease and psoriasis (*117*). A coding variant in glucokinase regulatory protein is associated with triglyceride levels and fasting glucose (*101*) but also with C reactive protein levels (*118*, *119*) and Crohn's disease (*86*). A SNP in TCF2 is associated with protection from T2D, as expected on the basis of Mendelian mutations at the same gene (*120*). Unexpectedly, the same association signal turned up in a GWAS for prostate cancer (*121*). Similarly, JAZF1 was identified as containing SNPs associated with T2D (*83*) and prostate cancer (*122*), and TCF7L2 with T2D (*123*) and colon cancer (*124*, *125*).

From Common SNPs to the Full Allelic Spectrum

The current HapMap provides reliable proxies for the vast majority of SNPs at frequencies above 5%, but its coverage declines rapidly for lower-frequency alleles (*37*). Such lower-frequency alleles may be particularly important: Alleles with strong deleterious effects are constrained by natural selection from becoming too common. We divide these alleles into two conceptually distinct classes:

1) Common variants with frequencies below 5%. By "common," we refer to variants that occur at sufficient frequency to be cataloged in studies of the general population and measured (directly, or indirectly through LD) in association studies. In practice, this class may include allele frequencies in the range of 0.5% and above. A GWAS of 2000 cases and 2000 controls provides good power for a 1% allele causing a factor of 4 increase in risk (even at $P < 10^{-8}$) (Fig. 2).

The value of lower-frequency common variants is illustrated by PCSK9 (proprotein convertase subtilisin/kexin type 9). The gene encoding PCSK9 contains very rare mutations causing autosomal dominant hypercholesterolemia (discovered by linkage analysis), as well as high-frequency common variants with modest effects. The former are too rare and the latter too weak to enable effective clinical study of PCSK9 with respect to coronary artery disease risk. Hobbs and Cohen sequenced the gene (*126*, *127*) and identified low-frequency common variants (0.5 to 1%), which allowed epidemiological research documenting a protective effect on myocardial infarction (*128*).

2) Rare variants. Most Mendelian diseases involve rare mutations that are essentially never observed in the general population. Rare mutations likely also play an important role in common diseases. Because they are numerous and individually rare, it is not possible to create a complete catalog in the general population. Instead, they must be identified by sequencing in cases and controls in each study. Moreover, because each variant is too rare to prove statistical evidence of association, the mutations must be aggregated as a class to compare the overall frequency of cases versus controls.

A few examples are known through candidate gene studies. Rare nonsynonymous mutations in MC4R are found in patients with extreme early-onset obesity (*129*). Rare nonsynonymous mutations in ABCA1 are more common in patients with extremely low HDL than in those with high HDL (*130*). An excess of rare mutations in renal salt-handling genes has been associated with lower blood pressure and protection against hypertension (*131*).

The sample size required to perform a genome-wide search based on coding mutations depends on the background frequency (μ) of mutations that confer disease risk and the level (ω) of increased risk for each such mutation. ABCA1 is a favorable case because μ and ω are high (the gene has an unusually large coding region of ~7 kb, and mutations confer a factor of ~6 increase in risk). Achieving genome-wide significance will likely require resequencing studies of thousands of cases and controls, similar to GWASs (Fig. 2).

GWASs of rare variants are already under way for large structural variants through the use of microarray analysis. A recent GWAS of autism revealed that a highly penetrant, recurrent microdeletion and microduplication of a 593-kb region in 16p11.2 explains 1% of cases (*132*). Moreover, several recent studies report that patients with autism and schizophrenia may have an excess of rare deletions across the genome relative to unaffected controls (*133*, *134*). Although these studies did not identify specific loci (none of the novel loci were observed more than once), they suggest that the universe of rare structural changes contributing to each disease may be as large and diverse as that of common SNPs.

The Genetic Architecture of Common Disease

Variants so far identified by GWASs together explain only a small fraction of the overall inherited risk of each disease (for example, ~10% of the variance for Crohn's and ~5% for T2D). Where is the remaining genetic variance to be found? There are several answers:

1) At disease loci already identified by GWAS, the locus-attributable risk will often be higher than currently estimated. This is because marker SNPs used in GWASs will typically be imperfect proxies for the actual causal mutation that led to the association signal. The causal gene will often contain additional mutations not tagged by the initial marker SNPs, both common and rare. Determining the contribution of each gene will require intensive studies of variants at each locus.

2) Many more disease loci remain to be identified by GWAS. As noted above, GWASs to date have had low statistical power and thus necessarily missed many loci with common variants of similar and smaller effects. The first studies did not have proxies for common structural variants and have failed to capture lower-frequency common variants (0.5 to 5%). Moreover, the vast majority of studies have been performed only in samples of European ancestry. Larger, more comprehensive, and more diverse GWASs will reveal many more loci.

3) Some disease loci will contain only rare variants. Such loci (if not already found by Mendelian genetics) cannot be identified by study of common variants alone. They will require systematic resequencing of all genes in large samples (Fig. 2).

4) Current estimates of the variance explained are based on simplifying assumptions. Because the genotype-phenotype correlation has yet to be well characterized, the estimates assume that the variants interact in a simple additive manner. Yet gene-gene and gene-environment interactions play important roles in disease risk. Although searches have not yet found much evidence for epistasis [e.g., (*93*, *94*, *135*)], this may simply reflect limited power to assess the many possible modes of interaction, including pairwise interactions and threshold effects. Once patterns of association and interaction are understood, effects of specific gene and environmental exposures on each phenotype may be larger.

For these reasons, it is premature to make inferences about the overall genetic architecture of common disease. Only by systematically exploring each of these directions over the coming years will a general picture emerge—with the likely

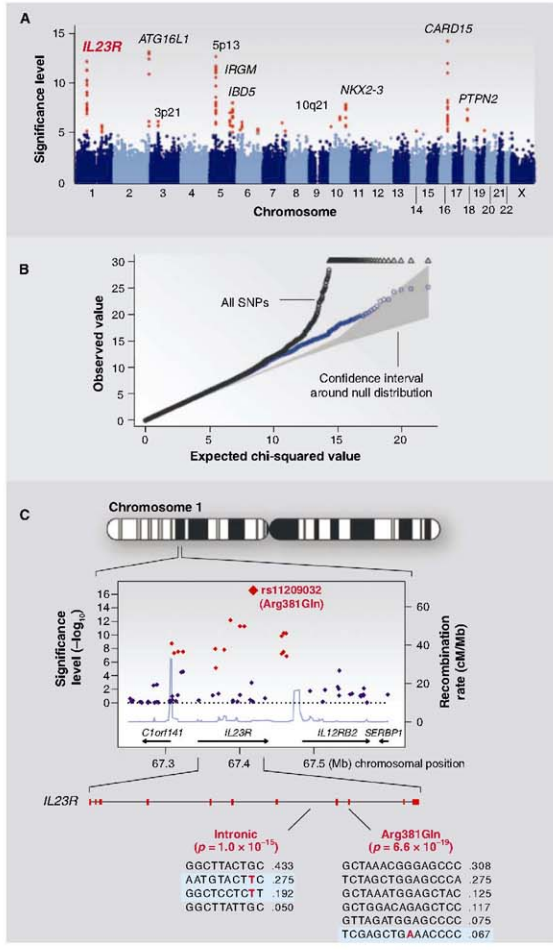


Fig. 3. GWAS for Crohn's disease. The panels show data from the study of Crohn's disease by the Wellcome Trust Case Control Consortium. **(A)** Significance level (P value on $\log_{10}$ scale) for each of the 500,000 SNPs tested across the genome. SNP locations reflect their positions across the 23 human chromosomes. SNPs with significance levels exceeding 10^{-5} (corresponding to 5 on the y axis) are colored red; the remaining SNPs are in blue. Ten regions with multiple significant SNPs are shown, labeled by their location or by the likely disease-related gene (e.g., *IL23R* on chromosome 1). **(B)** The fact that the SNPs in red are extreme outliers is made clear from a so-called Q-Q plot. A Q-Q plot is made as follows: The SNPs are ordered (from 1 to n) according to their observed P values; observed and expected P values are plotted for each SNP. Under the null distribution, the expected P value for the ith SNP is i/n. If there are no significant associations, the Q-Q plot will lie along the 45° line; the gray region corresponds to a 95% confidence region around this null expectation. Black points correspond to all 500,000 SNPs studied that passed strict quality control; they diverge strongly from the null expectation. Blue points reflect the P values that remain when the SNPs in the 10 most significant regions are removed; there is still some excess of significant P values, indicating the presence of additional loci of more modest effect. **(C)** Close-up of the region around the IL23R locus on chromosome 1. The first part shows the significance levels for SNPs in a region of ~400 kb, with colors as in (A). The highest significance level occurs at a SNP in the coding region of the *IL23R* gene (causing an $Arg^{381} \rightarrow Gln$ change). The light blue curve shows the inferred local rate of recombination across the region. There are two clear hotspots of recombination, with SNPs lying between these hotspots being strongly correlated in a few haplotypes. The second part shows that the IL23R locus harbors at least two independent, highly significant disease-associated alleles. The first site is the $Arg^{381} \rightarrow Gln$ polymorphism, which has a single disease-associated haplotype (shaded in blue) with frequency of 6.7%. The second site is in the intron between exons 7 and 8; it tags two disease-associated haplotypes with frequencies of 27.5% and 19.2%.

outcome being that different diseases will each be characterized by a different balance of allele frequencies, interactions, and types. Although the proportion of genetic variance explained is certain to grow in the coming years, it is unlikely to approach 100% because of practical limitations, such as the difficulty of detecting common variants with extremely small effects, genes harboring rare variants at very low frequency, and complex interactions among genes and with the environment.

Disease Risk Versus Disease Mechanism

The primary value of genetic mapping is not risk prediction, but providing novel insights about mechanisms of disease. Knowledge of disease pathways (not limited to the causal genes and mutations) can suggest strategies for prevention, diagnosis, and therapy. From this perspective, the frequency of a genetic variant is not related to the magnitude of its effect, nor to the potential clinical value that may be obtained.

The classic example is Brown and Goldstein's studies of FH, which affects ~0.2% of the population and accounts for a tiny fraction of the heritability of LDL and myocardial infarction. Studies of FH led

to the discovery of the LDL receptor and supported the development of HMGCR inhibitors (statins) for lowering LDL, the use of which is not limited to FH carriers.

More recently, GWASs have shown that common genetic variation in LDLR and HMGCR influences LDL levels (*84*, *85*). Although SNPs in HMGCR have only a small effect (~5%) on LDL levels, drugs targeting the encoded protein decrease LDL levels by a much greater extent (~30%). This is because the effect of an inherited variant is limited by natural selection and pleiotropy, whereas the effect of a drug treatment is not.

The Path Ahead

Given the long-standing success of genetic mapping in providing new insights into biology and disease etiology, and the recent proof that systematic association studies can identify novel loci, our aim should be nothing less than identifying all pathways at which genetic variation contributes to common diseases. We sketch key steps in achieving this goal.

Expanding clinical studies. Current studies are underpowered for the types of SNP alleles that we now know exist, and available evidence indicates that increasing sample size will yield substantial returns. A study of 1000 cases and 1000 controls provides only 1% power to detect a 20% variant that increases risk by a factor of 1.3, but a study of 5000 cases and 5000 controls provides 98% power (Fig. 2). Moreover, early data on rare single-nucleotide (*130*, *131*) and structural variants (*133*, *134*, *136*) indicate that similarly large samples will be needed to achieve the levels of statistical significance required to detect rare events in a genome-wide search.

Nearly all GWASs to date have been performed in populations of European ancestry. Even if a variant has the same effect in all ancestry groups, it may be more readily detected in one population simply because it happens to have higher frequency. Genetic effects will likely vary across groups because of modification by environment and behavior, which may vary more across groups than does genotype.

Many important diseases remain to be studied by GWAS. Disease-related intermediate traits can also offer substantial insight, particularly in conjunction with clinical endpoints. For example, newly described variants on chromosome 1 (near SORT1) are associated both with levels of LDL cholesterol (*84*, *85*) and with risk of myocardial infarction (*106*); this provides not only increased statistical confidence, but also a biomarker for gene function and pathophysiological insight. Genetic variants that influence gene expression [e.g., (*137*)] hold promise for elucidating regulatory pathways. Mapping of modifiers of Mendelian mutations—for example, genes that influence the age of onset in carriers of BRCA1 and BRCA2 mutations—may suggest ways to reverse high risk due to mutations.

Correlations between genetic variants and phenotypes are limited by the accuracy with which each is measured. The ability to measure genotype now far exceeds our ability to measure phenotype. Continuous ambulatory monitoring, imaging methods, and comprehenive ("-omic") approaches to biological samples all have promise in improving the accuracy of phenotype measurement.

Environmental exposures play a larger role in human phenotypic variation than does genetic variation, but environmental exposures are fundamentally more difficult to measure. DNA is stable throughout life, with a single physical chemistry that enables generic approaches to measurement. Environmental exposures are heterogeneous and may be fleeting. Improved methods for measuring environmental exposures, perhaps based on epigenetic marks they leave, are sorely needed.

Expanding the range of genetic variation. The lowest-hanging fruit will be to resequence loci that have been definitively implicated in disease by Mendelian genetics or by GWAS. Because the prior probability of a true association is higher, such regions will be the best setting to develop methods for understanding the statistical significance and biological importance of rare mutations. Initially, resequencing of coding exons will be easiest to interpret. Rare coding mutations with large effect will be especially valuable, because physiological studies of mutation carriers can help illuminate the biological basis of the disease, and because coding mutations of large effect are more straightforwardly transferred to cellular and animal models for mechanistic studies.

Extending GWASs to include structural variants and lower-frequency common variants will require comprehensive catalogs of genomic variation, as well as characterization of LD relationships. With new massively parallel sequencing technologies, an accurate map of all 1% alleles (both single-nucleotide and structural) should be achievable. A "1000 Genomes Project" was recently launched toward this end (*138*).

Some loci may harbor neither common variants nor rare structural variants, and thus will be missed by array and LD-based approaches. Discovering such genes will require sequencing in thousands of cases and controls. Initial studies will likely focus on exons, where functional mutations are enriched to the greatest extent. Highly parallel methods to capture hundreds of thousands of exons, and other targets of interest, are under development (*139*).

Multiple instances of de novo coding mutations at a locus (by comparing affected individuals with parents) could provide particularly powerful association information, because the human mutation rate is so low (in the range of 10^{-8}). But identifying de novo mutations without being overwhelmed by false positives will require extraordinary sequencing accuracy (far better than finished genome sequence). Because such studies will be expensive at first, priority should go to disorders with high heritability, where there is an unmet medical need, and for which other approaches have met with limited success. Psychiatric disorders might represent one such target.

Eventually, it will become practical to resequence entire genomes from thousands of cases and controls. The problem of interpretation will be much harder for noncoding functional elements, because it is unclear either how to aggregate elements to achieve a large enough target size, or to develop ways to recognize function-altering changes.

Routine genome sequencing of deeply phenotyped cohorts will fundamentally change the nature of genetic mapping: from the current serial process (in which initial localization by linkage or GWAS is followed by scrutiny of DNA variation and phenotypes) to a joint estimation procedure combining variation information of all types, frequencies, and phenotypes to discover and characterize genotype-phenotype correlations. New statistical methods will be required to combine evidence from rare and common alleles at a locus and across multiple loci, phenotypes, and nongenetic exposures. A particular challenge will be to identify mutations in regions without known function or evolutionary conservation.

There may be inherent limits to our ability to relate phenotypic variation and genotypic variation. To the extent that disease is influenced by tiny effects at hundreds of loci or highly heterogeneous rare mutations, it may be impractical to assemble sufficiently large samples to give a complete accounting.

Implications for Biology, Medicine, and Society

Genetic mapping is only a first step toward biological understanding and clinical application. Useful tools will include maps of evolutionary conservation (*108*) and chromatin state (*140*), as well as databases of cell-state signatures, such as genome-wide expression patterns, that may integrate aspects of cell biology under resting and provoked conditions (*141*). Creation of disease models, both in human cell culture and nonhuman animals, will be key. Physiological studies in patients classified by genotype may inform disease processes and lead to useful nongenetic biomarkers. Given the limits of human clinical research, rare alleles of strong effect may be more useful than common alleles of weak effect.

The high failure rate of clinical trials testifies to the limited predictive value of current approaches. By focusing attention on genes and processes, human genetics has the potential to yield productive targets and predictive animal models. In clinical trials, the ability to stratify patients by genotype or biological pathway may reveal differences in therapeutic response. Genetics may also increase the efficiency of outcome trials by focusing on patients at higher-than-average risk.

The extent to which genetic information will figure in "personalized medicine" will depend on whether predictive accuracy beyond conventional measures can be attained, and whether there are interventions whose effectiveness is improved by knowledge of a genetic test. Knowledge of a common variant that increases T2D

risk by 20% may eventually lead to new understanding and therapeutic strategies, but whether an increase in absolute risk (from 8% to 10%) is useful for patients remains to be seen. Although it is tempting to think that knowledge of individual risk might promote greater adherence to a healthy lifestyle, human behavior is complex and risk estimates are challenging to interpret. Even where genotype can predict response to a drug with a narrow therapeutic window, it cannot be assumed that genetic testing will necessarily lead to improved clinical outcomes.

Our understanding of complex disease will be in constant flux over the coming years. The pace of discovery, while scientifically exhilarating, poses daunting challenges. Direct-to-consumer marketing of genetic information is already under way. It will be a challenge for the public to understand the difference between relative and absolute risk, and to figure in their thinking the larger component of genetic and environmental factors not yet captured by today's technologies. Rigorous assessment of health benefit and cost are needed, including costs of testing and treatment that may flow from an altered sense of risk. As genetic information is shown to be useful, equitable access will be critical.

Finally, we must ensure that the promise of research on genetic factors in complex disease does not encourage a mistaken sense of genetic determinism. This is especially important for behavioral traits, which are especially prone to misinterpretation and misguided policy. We must constantly remind the public—and ourselves—that although genes play a role (and can lead us to new biological insight), our traits are powerfully shaped by the environment, and the solutions to important problems will often lie outside our genes.

References

1. A. Sturtevant, *J. Exp. Zool.* **14**, 43 (1913).
2. L. Clarke, J. Carbon, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2173 (1980).
3. W. Bender *et al.*, *Science* **221**, 23 (1983).
4. I. Aird, H. H. Bentall, J. A. Mehigan, J. A. Roberts, *Br. Med. J.* **2**, 315 (1954).
5. V. M. Ingram, *Nature* **178**, 792 (1956).
6. T. D. Petes, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5091 (1977).
7. A. J. Jeffreys, *Cell* **18**, 1 (1979).
8. Y. W. Kan, A. M. Dozy, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5631 (1978).
9. D. Botstein, R. L. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
10. J. F. Gusella *et al.*, *Nature* **306**, 234 (1983).
11. H. Donis-Keller *et al.*, *Cell* **51**, 319 (1987).
12. C. Dib *et al.*, *Nature* **380**, 152 (1996).
13. T. J. Hudson *et al.*, *Science* **270**, 1945 (1995).
14. Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/sites/entrez?db=omim).
15. P. L. Welcsh, M. C. King, *Hum. Mol. Genet.* **10**, 705 (2001).
16. R. P. Lifton, *Harvey Lect.* **100**, 71 (2004).
17. G. I. Bell, K. S. Polonsky, *Nature* **414**, 788 (2001).
18. E. East, *Am. Nat.* **44**, 65 (1910).
19. E. Altenburg, H. J. Muller, *Genetics* **5**, 1 (1920).
20. R. A. Fisher, *Trans. R. Soc. Edinburgh* **52**, 399 (1918).
21. A. H. Paterson *et al.*, *Nature* **335**, 721 (1988).
22. J. Klein, A. Sato, *N. Engl. J. Med.* **343**, 782 (2000).
23. W. J. Strittmatter, A. D. Roses, *Annu. Rev. Neurosci.* **19**, 53 (1996).
24. J. N. Hirschhorn, K. Lohmueller, E. Byrne, K. Hirschhorn, *Genet. Med.* **4**, 45 (2002).
25. K. E. Lohmueller, C. L. Pearce, M. Pike, E. S. Lander, J. N. Hirschhorn, *Nat. Genet.* **33**, 177 (2003).
26. F. S. Collins, M. S. Guyer, A. Chakravarti, *Science* **278**, 1580 (1997).
27. E. S. Lander, *Science* **274**, 536 (1996).
28. N. Risch, K. Merikangas, *Science* **273**, 1516 (1996).
29. M. Kimura, T. Ota, *Genetics* **75**, 199 (1973).
30. H. Harris, *Proc. R. Soc. London Ser. B* **164**, 298 (1966).
31. W. H. Li, L. A. Sadler, *Genetics* **129**, 513 (1991).
32. R. Sachidanandam *et al.*, *Nature* **409**, 928 (2001).
33. R. Lewontin, in *Evolutionary Biology 6*, T. Dobzhansky, M. K. Hecht, W. C. Steere, Eds. (Appleton-Century-Crofts, New York, 1972), pp. 391–398.
34. D. E. Reich, E. S. Lander, *Trends Genet.* **17**, 502 (2001).
35. D. G. Wang *et al.*, *Science* **280**, 1077 (1998).
36. Entrez SNP (www.ncbi.nlm.nih.gov/sites/entrez?db=snp).
37. International HapMap Consortium, *Nature* **437**, 1299 (2005).
38. S. E. Antonarakis, C. D. Boehm, P. J. Giardina, H. H. Kazazian Jr., *Proc. Natl. Acad. Sci. U.S.A.* **79**, 137 (1982).
39. E. S. Lander, D. Botstein, *Cold Spring Harb. Symp. Quant. Biol.* **51**, 49 (1986).
40. B. Kerem *et al.*, *Science* **245**, 1073 (1989).
41. J. Hastbacka *et al.*, *Nat. Genet.* **2**, 204 (1992).
42. L. Kruglyak, *Nat. Genet.* **22**, 139 (1999).
43. K. G. Ardlie, L. Kruglyak, M. Seielstad, *Nat. Rev. Genet.* **3**, 299 (2002).
44. M. J. Daly, J. D. Rioux, S. F. Schaffner, T. J. Hudson, E. S. Lander, *Nat. Genet.* **29**, 229 (2001).
45. N. Patil *et al.*, *Science* **294**, 1719 (2001).
46. S. B. Gabriel *et al.*, *Science* **296**, 2225 (2002); published online 23 May 2002 (10.1126/science.1069424).
47. G. C. Johnson *et al.*, *Nat. Genet.* **29**, 233 (2001).
48. D. E. Reich *et al.*, *Nat. Genet.* **32**, 135 (2002).
49. D. C. Crawford *et al.*, *Nat. Genet.* **36**, 700 (2004).
50. G. A. T. McVean *et al.*, *Science* **304**, 581 (2004).
51. S. A. Tishkoff, B. C. Verrelli, *Annu. Rev. Genomics Hum. Genet.* **4**, 293 (2003).
52. International HapMap Consortium, *Nature* **449**, 851 (2007).
53. J. Sebat *et al.*, *Science* **305**, 525 (2004).
54. A. J. Iafrate *et al.*, *Nat. Genet.* **36**, 949 (2004).
55. E. Tuzun *et al.*, *Nat. Genet.* **37**, 727 (2005).
56. D. A. Hinds, A. P. Kloek, M. Jen, X. Chen, K. A. Frazer, *Nat. Genet.* **38**, 82 (2006).
57. S. A. McCarroll *et al.*, *Nat. Genet.* **38**, 86 (2006).
58. D. P. Locke *et al.*, *Am. J. Hum. Genet.* **79**, 275 (2006).
59. R. Redon *et al.*, *Nature* **444**, 444 (2006).
60. S. A. McCarroll *et al.*, *Nat. Genet.* **40**, 1166 (2008).
61. S. S. Deeb *et al.*, *Nat. Genet.* **20**, 284 (1998).
62. D. Altshuler *et al.*, *Nat. Genet.* **26**, 76 (2000).
63. J. C. Florez, J. N. Hirschhorn, D. Altshuler, *Annu. Rev. Genomics Hum. Genet.* **4**, 257 (2003).
64. I. Pe'er, R. Yelensky, D. Altshuler, M. J. Daly, *Genet. Epidemiol.* **32**, 381 (2008).
65. W. C. Knowler, R. C. Williams, D. J. Pettitt, A. G. Steinberg, *Am. J. Hum. Genet.* **43**, 520 (1988).
66. R. S. Spielman, R. E. McGinnis, W. J. Ewens, *Am. J. Hum. Genet.* **52**, 506 (1993).
67. B. Devlin, K. Roeder, *Biometrics* **55**, 997 (1999).
68. J. K. Pritchard, N. A. Rosenberg, *Am. J. Hum. Genet.* **65**, 220 (1999).
69. A. L. Price *et al.*, *Nat. Genet.* **38**, 904 (2006).
70. D. G. Clayton *et al.*, *Nat. Genet.* **37**, 1243 (2005).
71. J. M. Chapman, J. D. Cooper, J. A. Todd, D. G. Clayton, *Hum. Hered.* **56**, 18 (2003).
72. D. A. Hinds *et al.*, *Science* **307**, 1072 (2005).
73. P. I. W. de Bakker *et al.*, *Nat. Genet.* **37**, 1217 (2005).
74. J. Marchini, B. Howie, S. Myers, G. McVean, P. Donnelly, *Nat. Genet.* **39**, 906 (2007).
75. K. M. Weiss, J. D. Terwilliger, *Nat. Genet.* **26**, 151 (2000).
76. J. Couzin, *Science* **296**, 1391 (2002).
77. R. J. Klein *et al.*, *Science* **308**, 385 (2005); published online 10 March 2005 (10.1126/science.1109557).
78. A. O. Edwards *et al.*, *Science* **308**, 421 (2005); published online 10 March 2005 (10.1126/science.1110189).
79. J. L. Haines *et al.*, *Science* **308**, 419 (2005); published online 10 March 2005 (10.1126/science.1110359).
80. G. Thorleifsson *et al.*, *Science* **317**, 1397 (2007); published online 9 August 2007 (10.1126/science.1146554).
81. Wellcome Trust Case Control Consortium, *Nature* **447**, 661 (2007).
82. D. Altshuler, M. Daly, *Nat. Genet.* **39**, 813 (2007).
83. E. Zeggini *et al.*, *Nat. Genet.* **40**, 638 (2008).
84. S. Kathiresan *et al.*, *Nat. Genet.* **40**, 189 (2008).
85. C. J. Willer *et al.*, *Nat. Genet.* **40**, 161 (2008).
86. J. C. Barrett *et al.*, *Nat. Genet.* **40**, 955 (2008).
87. G. Lettre *et al.*, *Nat. Genet.* **40**, 584 (2008).
88. M. N. Weedon *et al.*, *Nat. Genet.* **40**, 575 (2008).
89. S. Sanna *et al.*, *Nat. Genet.* **40**, 198 (2008).
90. D. F. Gudbjartsson *et al.*, *Nat. Genet.* **40**, 609 (2008).
91. S. A. McCarroll *et al.*, *Nat. Genet.* **40**, 1107 (2008).
92. C. A. Haiman *et al.*, *Nat. Genet.* **39**, 638 (2007).
93. J. Maller *et al.*, *Nat. Genet.* **38**, 1055 (2006).
94. M. Li *et al.*, *Nat. Genet.* **38**, 1049 (2006).
95. R. R. Graham *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6758 (2007).
96. R. H. Duerr *et al.*, *Science* **314**, 1461 (2006); published online 26 October 2006 (10.1126/science.1135245).
97. R. Sladek *et al.*, *Nature* **445**, 881 (2007).
98. D. F. Easton *et al.*, *Nature* **447**, 1087 (2007).
99. E. Zeggini *et al.*, *Science* **316**, 1336 (2007); published online 25 April 2007 (10.1126/science.1142364).
100. L. J. Scott *et al.*, *Science* **316**, 1341 (2007); published online 25 April 2007 (10.1126/science.1142382).
101. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes for BioMedical Research, *Science* **316**, 1331 (2007); published online 26 April 2007 (10.1126/science.1142358).
102. A. Rivera *et al.*, *Hum. Mol. Genet.* **14**, 3227 (2005).
103. L. T. Amundadottir *et al.*, *Nat. Genet.* **38**, 652 (2006).
104. M. L. Freedman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14068 (2006).
105. R. McPherson *et al.*, *Science* **316**, 1488 (2007); published online 2 May 2007 (10.1126/science.1142447).
106. N. J. Samani *et al.*, *N. Engl. J. Med.* **357**, 443 (2007).
107. A. Helgadottir *et al.*, *Science* **316**, 1491 (2007); published online 2 May 2007 (10.1126/science.1142842).
108. R. H. Waterston *et al.*, *Nature* **420**, 520 (2002).
109. P. R. Burton *et al.*, *Nat. Genet.* **39**, 1329 (2007).
110. M. Cargill *et al.*, *Am. J. Hum. Genet.* **80**, 273 (2007).
111. J. C. Florez *et al.*, *N. Engl. J. Med.* **355**, 241 (2006).
112. N. Grarup *et al.*, *Diabetes* **56**, 3105 (2007).
113. L. Pascoe *et al.*, *Diabetes* **56**, 3101 (2007).
114. R. Saxena *et al.*, *Diabetes* **55**, 2890 (2006).
115. H. Staiger *et al.*, *PLoS One* **2**, e832 (2007).
116. V. Steinthorsdottir *et al.*, *Nat. Genet.* **39**, 770 (2007).
117. N. Wolf *et al.*, *J. Hum. Genet.* **45**, 114 (2008).
118. A. P. Reiner *et al.*, *Am. J. Hum. Genet.* **82**, 1193 (2008).
119. P. M. Ridker *et al.*, *Am. J. Hum. Genet.* **82**, 1185 (2008).
120. W. Winckler *et al.*, *Diabetes* **56**, 685 (2007).
121. J. Gudmundsson *et al.*, *Nat. Genet.* **39**, 977 (2007).
122. G. Thomas *et al.*, *Nat. Genet.* **40**, 310 (2008).
123. S. F. A. Grant *et al.*, *Nat. Genet.* **38**, 320 (2006).
124. A. Hazra *et al.*, *Cancer Causes Control* **19**, 975 (2008).
125. A. R. Folsom *et al.*, *Diabetes Care* **31**, 905 (2008).
126. I. K. Kotowski *et al.*, *Am. J. Hum. Genet.* **78**, 410 (2006).
127. J. Cohen *et al.*, *Nat. Genet.* **37**, 161 (2005).
128. J. C. Cohen, E. Boerwinkle, T. H. Mosley Jr., H. H. Hobbs, *N. Engl. J. Med.* **354**, 1264 (2006).
129. J. N. Hirschhorn, D. Altshuler, *J. Clin. Endocrinol. Metab.* **87**, 4438 (2002).
130. J. C. Cohen *et al.*, *Science* **305**, 869 (2004).
131. W. Ji *et al.*, *Nat. Genet.* **40**, 592 (2008).
132. L. A. Weiss *et al.*, *N. Engl. J. Med.* **358**, 667 (2008).
133. T. Walsh *et al.*, *Science* **320**, 539 (2008); published online 27 March 2008 (10.1126/science.1155174).
134. J. Sebat *et al.*, *Science* **316**, 445 (2007); published online 14 March 2007 (10.1126/science.1138659).
135. J. D. Rioux *et al.*, *Nat. Genet.* **39**, 596 (2007).
136. L. A. Weiss *et al.*, *N. Engl. J. Med.* **358**, 667 (2008).
137. V. Emilsson *et al.*, *Nature* **452**, 423 (2008).
138. 1000 Genomes (www.1000genomes.org).
139. T. J. Albert *et al.*, *Nat. Methods* **4**, 903 (2007).
140. T. S. Mikkelsen *et al.*, *Nature* **448**, 553 (2007).
141. J. Lamb *et al.*, *Science* **313**, 1929 (2006).

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/881/DC1
Table S1

10.1126/science.1156409

INTRODUCTION

From Genes to Social Behavior

WHEN IT COMES TO BEHAVIOR, WE HAVE MOVED BEYOND GENETIC DETERMINISM. Our genes do not lock us into certain ways of acting; rather, genetic influences are complicated and mutable and are only one of many factors affecting behavior. In their editorial, Landis and Insel (p. 821) elaborate on this idea, explaining that proteins encoded by genes direct the formation of multicomponent neural circuits, which are the true substrates of behavior, as these circuits respond to internal and outside stimuli.

Why do we study the genetic underpinnings of behavior? One reason is to understand how certain behaviors evolve. Conserved neural pathways can be tied to the evolution of social behaviors (Robinson *et al.*, p. 896), and the conserved peptides oxytocin and vasopressin regulate social cognition and reproductive behaviors in many species (Donaldson and Young, p. 900). In a News story, Pennisi focuses on a region of chromosome 17 that has a complicated pattern of evolution in humans and other primates and is linked in unexpected ways to various disorders, including mental retardation, learning disabilities, and dementias.

Genetics can help us understand why identical circumstances can elicit different behavioral responses among individuals. Genetic differences are reflected in variations in behavior; activation of distinct versions of a hormone receptor gene, an example Donaldson and Young present, results in monogamous behavior in one species of vole but not in another. Conversely, as Robinson *et al.* describe, insights from recent work show that perceiving social information—such as bird songs or dominance behavior from cichlid fish—from another individual of the same species can itself alter gene expression in the brain, with downstream effects on physiology and behavior.

The potent genetic tools available for *Drosophila* have uncovered many genes that, when deleted, disrupt behaviors. This, in turn, has allowed dissection of the neural circuits that control essential behaviors. One of the best understood is a social activity necessary for reproduction—stereotypical mating behavior—as outlined by Dickson (p. 904). Genetic methods have also led to the understanding of another class of behaviors: those driven by the circadian clock. The genetic basis of the clock was elegantly worked out in *Drosophila*, followed by a similar achievement in mice. The reasons for these successes are outlined by Takahashi in his Perspective (p. 909), in which he also explains what tools will be needed to attain similar advances for other behaviors in mice.

Humans are not as genetically tractable as mice or flies, and human behavior is not as stereotypical. Holden's News story on the strengths and shortcomings of genetic studies of personality illustrates this point (p. 892). So do Cotton and some members of the Human Variome Project community in a Policy Forum (p. 861) that describes how the genes and loci associated with disorders of the nervous system are a particular challenge to geneticists and clinical neurologists in need of reliable diagnostic tests. And in a Perspective on a critical human social activity—politics—Fowler and Schreiber (p. 912) argue that genetics and neurobiology have much to teach us about how our leaders are chosen.

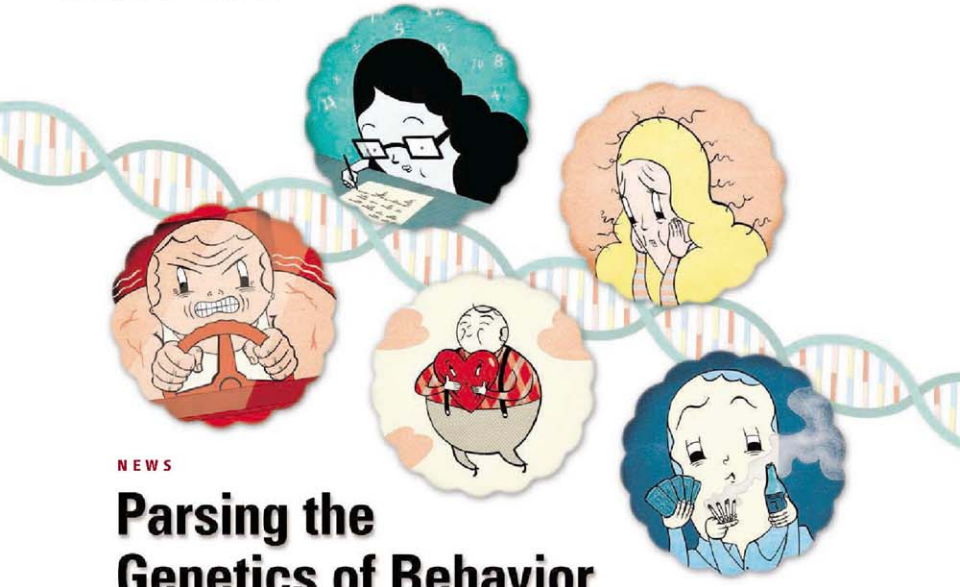
Some believe that psychology is the last frontier of genetic analysis. This special section provides a sampling of our early explorations.

–BARBARA R. JASNY, KATRINA L. KELNER, ELIZABETH PENNISI

Genetics of Behavior

CONTENTS

See also related Editorial on p. 821, News story by Pennisi, Policy Forum on p. 861, and Science Podcast at www.sciencemag.org/multimedia/podcast



NEWS

Parsing the Genetics of Behavior

It takes more than one gene, or even a few genes, to make a personality trait. But which ones?

SOME YEARS AGO, A SCIENTIST-EDUCATOR told *Science* she would never be convinced of a biological basis for sex differences in math performance until someone showed her a "math gene." The comment rests on a commonly held misconception: that simple one-to-one links exist between a gene and each facet of our personalities. Headlines such as " 'Ruthlessness' Gene Discovered" or " 'Divorce Gene' Linked to Relationship Troubles" feed that impression.

For some of us, it's satisfying to attribute social awkwardness to anxiety genes or to think that the driver who cuts off other cars as he zips across lanes is pumped up by the "warrior" gene. Was it a bad dopamine receptor gene that made author Ernest Hemingway prone to depression? Can variations in a vasopressin receptor gene—a key to monogamy in voles—help explain adulterous behavior?

But as scientists are discovering, nailing down the genes that underlie our unique personalities has proven exceedingly difficult. That genes strongly influence how we act is beyond question. Several decades of twin, family, and adoption studies have demonstrated that roughly half of the variation in most behavioral traits can be chalked up to genetics. But identifying the causal chain in single-gene disorders such as Huntington's disease is child's play compared with the challenges of tracking genes contributing to, say, verbal fluency, outgoingness, or spiritual leanings. In fact, says Wendy Johnson, a psychologist at the University of Edinburgh, U.K., understanding genetic mechanisms for personality traits "is one of the biggest mysteries facing the behavioral sciences."

All we really know so far is that behavioral genes are not solo players; it takes many to orchestrate each trait. Complicating matters further, any single gene may play a role in several seemingly disparate functions. For example, the same gene may influence propensities toward depression, overeating, and impulsive behavior, making it difficult to tease out underlying mechanisms.

Each gene comes in a variety of flavors, or alleles, with varying degrees of sequence variation. One allele might contribute to a winning personality whereas another may raise the risk of mental illness. Environment plays a strong hand, bringing out, neutralizing, or even negating a gene's influence. And genes interact with one another in unpredictable ways.

Science took a look at a few genes that have been in the news, with an eye toward understanding just what we do—and can—know about genes behind individual variation in temperament and personality.

CREDIT: RYAN SNOOK; JESSICA NEWFIELD/SCIENCE

Loves me, loves me not …

A genetic screen for marital success? It sounds like a *Saturday Night Live* skit, but one Canadian company is actually offering just that sort of test. For $99, Genesis Biolabs in St. John's, Newfoundland, will examine your—or your partner's—*vasopressin 1a receptor* (*AVPR1a*) gene, which this year grabbed headlines once as the "ruthlessness gene" and again as a "divorce gene." But is the test really any more predictive than pulling petals off a daisy?

Vasopressin is a hormone involved in attachment to mates and offspring. Among voles, prairie voles are true to their mates. Meadow and montane voles prefer to play the field. Prairie voles have a few extra bases in the DNA in front of this gene, which influences how much and where vasopressin is released in the brain. This difference matters: Extra *AVPR1a* in the brain makes promiscuous meadow voles act more like monogamous prairie voles, spending more time with partners and grooming offspring (see p. 900).

Subsequent research has disproved any simple relationship between this gene and animal mating patterns. Nonetheless, scientists have now observed hints that variation in the human *AVPR1a* gene may influence the far more complex arena of human behavior.

A team led by Hasse Walum of the Karolinska Institute in Stockholm looked at the DNA preceding the *AVPR1a* gene in about 500 pairs of adult same-sex Swedish twins, all of them married or cohabiting for at least 5 years, and their partners. One short variant of a stretch of DNA in this region—there are several variants—was associated with less stable relationships. Answers to questions such as "How often do you kiss your mate?" and "How often are you and your partner involved in common interests outside the family?" reflected slightly lower feelings of attachment on the part of men with this variant, researchers reported in the 16 September issue of the *Proceedings of the National Academy of Sciences*. These men were less likely to be married and, among those in relationships, more likely to have experienced recent marital strife.

A gene worth testing to be assured of marital bliss? Not quite. "This is a brand-new study for which replication has not been attempted," Johnson points out.

Another paper published last spring showed a different link between *AVPR1a* and how people treat others. Richard Ebstein and colleagues at Hebrew University in Jerusalem reported that the length of the variant predicted how human subjects would respond in the "dictator game," a way to assess altruism. The researchers divided 200 volunteers into groups "A" and "B." The "A's" received $14 each and were told to share as much as they wished with a "B" whom they had never met. About 18% kept all the money, and 6% gave it all away, with the rest somewhere in the middle. The people who behaved more selfishly—or, as the headlines proclaimed, more ruthlessly—had the same variant as the people with the less stable relationships in the study mentioned earlier. Ebstein speculates that in these people, vasopressin receptors were distributed in such a way that they provided less of a sense of reward from the act of giving (or loving). He and other scientists suspect that short variants of this gene will be implicated in autism and related disorders, because a core feature of autism is the inability to make connections with other people.

Although such theories are intellectually appealing, there are few replicated studies to give them heft, notes psychologist Simon Easteal of Australian National University in Canberra. Too often, the subjects assessed are too different—How does one compare adolescents with married couples?—and the effect of environment too difficult to control for. So, getting reliable replications of studies involving behavior is, he says, "much harder than for studies of medical conditions."

A "bounce-back" gene

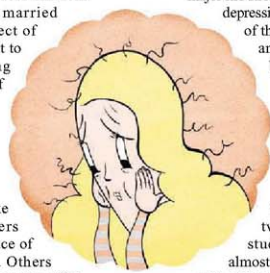
Some people are like Woody Allen characters who melt down in the face of the smallest obstacles. Others seem to have a thick hide against life's slings and arrows. The roots of such resilience may lie in a gene for a protein that regulates serotonin, a brain messenger that has been associated with emotional ups and downs. The gene is called *SERT* for serotonin transporter.

In a classic paper published in *Science* in 1996, Klaus-Peter Lesch of the University of Würzburg, Germany, and colleagues at the U.S. National Institutes of Health reported that the length of the regulatory DNA at the beginning of *SERT* affected human behavior. Lesch's team found that among 505 adults, those scoring high on various tests measuring "neuroticism"—depression and anxiety—tended to have one or two copies of a short variant whereas those who were more laid back had only the long form. The short version translates into more serotonin in the synapse, and too much serotonin leads to anxiety, in both animals and humans.

The short version accounted for up to 4% of the increase in anxiety and negative emotions in this group. Four percent doesn't sound like much, but it's huge for any personality trait, says psychologist Turhan Canli of Stony Brook University in New York state. In fact, he says, scientists have been able to find "no gene in the intervening years that has accounted for that much variability."

In another landmark study published in 2003, researchers showed that the effect of the gene depends on life experiences. In Dunedin, New Zealand, researchers led by Avshalom Caspi of the Institute of Psychiatry in London tracked 847 people over more than 20 years from the age of 3. The researchers counted stressful life events occurring between the ages of 21 and 26 and asked subjects if they had been depressed in the past year.

Among people who had not reported any major life stresses, the probability of depression was low regardless of their *SERT* alleles. But among people who had been through four or more stressful experiences, 43% of those with two short alleles reported a major depressive episode—more than double the proportion of subjects with two long alleles. The study also showed that almost two-thirds of people with a history of abuse as children experienced major depression as adults if they had two short alleles. But child abuse didn't raise the risk of adult depression in people with two long alleles.



CREDIT: RYAN SNOOK

Wanted: Math Gene

Last year, researchers at Washington University in St. Louis, Missouri, reported in the journal *Behavior Genetics* that certain aspects of IQ seemed to be related to *CHRM2* (*cholinergic muscarinic 2 receptor*), a gene whose protein is involved in pathways related to learning, memory, and problem-solving.

There, a team led by psychiatric geneticist Danielle Dick analyzed DNA and IQ test results from members of 200 families, 2150 individuals in all, as part of the Collaborative Study on the Genetics of Alcoholism. The team found a modest correlation between spatial and logical reasoning skills and certain variations in this gene.

But this study is one of very few to find any connection between genes and IQ—and it has yet to be replicated. This situation reflects a major paradox. Cognitive abilities are among the most genetically influenced of human behavioral traits: In studies over the years, scientists have estimated that somewhere between 40% to 80% of the variation in individual IQ scores in a given population is attributable to individual genetic differences. This is comparable to the genetic influence on height. Yet IQ genes have so far been impossible to nail down.

Psychologist Robert Plomin of the Institute of Psychiatry in London has spent years scouring genomes for signs of loci associated with high IQ. The largest study yet was a genome-wide scan of DNA from 6000 children using 500,000 markers that could help pinpoint relevant DNA. The study compared groups of low-IQ children with groups of high-IQ children in hopes of teasing out markers linked to intelligence.

A handful of markers had a significant association with the aspects of IQ deemed most heritable, such as verbal ability. But none accounted for more than 0.4% of the variance. In other words, if the IQs of the population in question ranged from 80 to 130 points, the biggest gene effect the researchers could find would account for less than one-quarter of an IQ point.

It seems to be much easier to identify genes for disabilities than for abilities. "The only genes we have identified so far for cognitive ability are for mental retardation, and there are about 300 of them," some of which have quite severe repercussions, says Wendy Johnson of the University of Edinburgh, U.K. Many are also associated with other types of disabilities. But corresponding genius-type alleles, particularly for specific skills such as math ability, don't seem to exist. **–C.H.**

Unfortunately, however, the picture is still unclear. Psychologists at the University of Bristol in the United Kingdom published a meta-analysis of studies of this gene in July in *Biological Psychiatry*. They concluded that the published studies weren't based on large enough samples and that the interaction effect between the gene and stressful life events is probably "negligible."

The more researchers look into this gene, the more widespread its associations appear to be, adding to the confusion. "The serotonin transporter is implicated in everything from heart disease to sleep disorders and irritable bowel syndrome [as well as] schizophrenia, depression, attention deficit hyperactivity disorder, autism, and sensation-seeking, to name just a few," says Johnson. With such a broad scope, its effects on behavior must be "extremely general," she notes. So to call it a resilience gene doesn't really fit.

Warrior gene

In 2006, a New Zealand researcher, Rod Lea, stirred up a political storm when he reported that a variant of a gene for monoamine oxidase-A (MAO-A)—which breaks down neurotransmitters—could be behind risk-taking and aggressive behavior in Māori, the indigenous people of New Zealand. The Māori have a warlike heritage, and a large proportion of this ethnic group carry a version of the gene shown in animal studies to be connected to aggressive behavior. Lea, a genetic epidemiologist at the Institute of Environmental Science and Research in Wellington, suggested that the gene could help explain Māori social and health problems such as fighting, gambling, and addictions. Although it's true that 60% of Asians (including Māori) carry the "warrior" variant (compared with 40% of Caucasians), Lea's critics quickly pointed out that it was too big a leap to ascribe Māori social problems to a single gene.

Yet brain-imaging studies "underscore the role of MAO-A [quite] specifically" in male aggressiveness, says neuroscientist Nelly Alia-Klein of Brookhaven National Laboratory in Upton, New York: Researchers have not detected connections between brain MAO-A and any other personality trait, she notes. In one study using functional magnetic resonance imaging, Andreas Meyer-Lindenberg and colleagues at the National Institute of Mental Health (NIMH) in Bethesda, Maryland, presented normal subjects with neutral or "emotionally aversive" images such as fearful faces. Monitoring activity in key brain regions such as the amygdala, the seat of fear, they found that the amygdalas of subjects with the "warrior" variant were hyperresponsive to such images. This sensitivity suggests that these individuals had problems regulating their emotions, which would also make them more likely to act on aggressive impulses, Meyer-Lindenberg reported.

But the gene variant isn't all that matters. Caspi's Dunedin study has shown that the environment—in the form of traumatic life events—plays a critical role in how this gene is expressed. Caspi's group reported in 2002 that the warrior *MAOA* variant is associated with violent and antisocial behavior but only in people with a history of abuse as children. These men were 2.8 times as likely as nonabused males with this genotype to develop behavioral problems that are often the precursor to a life of crime and drug abuse. Children with a different variant were less likely to develop antisocial problems in response to maltreatment (*Science*, 2 August 2002, p. 851).

Earlier this year, researchers drew similar conclusions based on the University of North Carolina's (UNC's) long-running National Longitudinal Study of Adolescent Health (NLSAH). Guang Guo of UNC Chapel Hill and colleagues analyzed genetic and social data from 1100 males and found that the undesirable effects from the "warrior" allele were only manifested when "social controls"—the steadying influence of a healthy

CREDIT: RYAN SNOOK

family and social environment—were absent. They reported these results in the August 2008 issue of the *American Sociological Review*.

Testosterone seems to add toxicity to the mix. Rickard Sjöberg of Uppsala University in Sweden and David Goldman of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in Bethesda compared the genes and testosterone levels of 95 male alcoholics who have criminal records with those of 45 nonalcoholic, law-abiding controls. They reported that the combination of low MAO-A and high testosterone spells antisocial behavior, as revealed by answers on an aggression scale. If these findings are replicated, Goldman says, they might help clear up the relationship of testosterone to aggression: Maybe the hormone causes trouble only in males who also have this gene variant, he says.

The warrior gene as the root of social ills may be dead, but it still has a fighting chance as a gene important to behavior.

Can't get no satisfaction

What do Janis Joplin, Amy Winehouse, and Jimi Hendrix have in common? If you want to find examples of people whose brain reward circuits have gone haywire, the world of rock stars is probably a good place to look. But is a dopamine receptor gene at the heart of these musicians' addictions?

Scientists have proposed that deficiencies in the brain messenger dopamine lead to various unhealthy forms of sensation-seeking to compensate for the failure to get a charge out of things that give most people pleasure. For years, they've been trying to nail down the role of dopamine receptors, in particular one called the D2 dopamine receptor, in addictions to alcohol, drugs, smoking, or gambling, as well as eating disorders and obesity.

The A1 allele of this gene yields receptors that don't work as well, and that translates into less dopamine firing up the reward circuits. Some scientists think this can lead to a tendency to abuse drugs and engage in impulsive, sensation-seeking, or antisocial behavior—including problems forming relationships.

Scientists led by anthropologist Dan Eisenberg of Northwestern University in Evanston, Illinois, reported last year in *Evolutionary Psychology* that in a group of 195 student subjects at Binghamton University in New York state, those with A1 alleles were more likely to engage in early sexual activity but were less inclined to develop steady relationships. This putative role in attachment has attracted the attention of political scientists looking for possible biological foundations for political behavior (see p. 912).

James Fowler and colleagues at the University of California, San Diego, picked up on reports such as this, as well as on animal research suggesting a connection between low dopamine receptor binding and low social bonding. They hypothesized that people with more efficient receptors—that is, with one or more A2 alleles—would be more trusting and therefore more likely to join a political party. After delving into NLSAH, they reported that, indeed, people with two A2 alleles (and no A1) were 8% more likely to form political attachments. Fowler called it "the first gene ever associated with partisan attachment."

But that's only the latest in the long and contradiction-riddled history of research on the D2 dopamine receptor gene. Guo looked for a link between social behavior and this gene by assessing delinquency rates in teenagers. What he found was that boys with one A1 allele tended to have higher delinquency rates than those with two copies of the A2 allele. But the rates were also higher than in those boys with two A1 copies, suggesting that there is not a simple relationship between the amount of dopamine and behavior. Warns Goldman: "There is still more heat than light with this gene."

CREDIT: RYAN SNOOK

Titrating anxiety

Scientists aren't doing much better at understanding the biological role of another player in the dopamine circuit. Dozens of studies have tried to figure out the gene for catechol *O*-methyltransferase (COMT), an enzyme that breaks down dopamine in the prefrontal cortex, the seat of higher cognitive functions such as planning and reasoning.

The two major variants of the gene code for enzymes that differ by one amino acid: The substitution of a valine for a methionine revs up the protein's activity fourfold. Both the high- and low-activity versions of the gene have their costs and benefits. Mice with the high-activity *COMT*—meaning less dopamine in the synapses—have poor memories and reduced sensitivity to pain. With the gene knocked out, and thus higher dopamine activity, mice show increased startle and anxiety responses.

In humans as well, different versions of the gene have been implicated in cognitive and emotional dysfunction, says Goldman. In several studies, people with two low-activity *COMT* genes have tested high for fear, anxiety, and negative thinking. A study at Yale University in 2005, for example, gave 497 undergraduates personality tests and found that those with low-activity *COMT* genes were more neurotic and less extraverted.

In research getting closer to the interface between biology and behavior, published in the August issue of *Behavioral Neuroscience*, researchers reported a difference in a simple test that has come to be recognized as a reliable indicator for anxiety: the startle reflex, as manifested in involuntary eye blinking in response to a sudden noise or unpleasant pictures. Among 96 female psychology students, individuals with two copies of the low-activity *COMT* had the most exaggerated startle responses, says Christian Montag of the University of Bonn in Germany.

Yet other work evaluating how well individuals organize their thoughts found low-activity *COMT* to be an asset. Psychiatrist Daniel Weinberger and colleagues at NIMH think they know why. Brain-imaging studies of 100 normal adults found that those with the low-activity *COMT* have denser nerve connections. Weinberger and others speculate that the elevated dopamine in the prefrontal cortex may bolster temporary connections, leading to better concentration but reduced ability to shift focus and more behavioral rigidity. As a result, a person may dwell excessively on stressful thoughts. So the gene seems to come with a tradeoff—better cognitive function but more anxiety—the scientists conclude.

The trouble with a lot of research on *COMT*, however, is that some studies find significant linkages only in women, and others don't find any at all. "*COMT* leaves a trail of intriguing hints," says Edinburgh's Johnson, "but nothing that solidly replicates."

–CONSTANCE HOLDEN

REVIEW

Genes and Social Behavior

Gene E. Robinson,[1,2,3]* Russell D. Fernald,[4] David F. Clayton[2,3,5]

What genes and regulatory sequences contribute to the organization and functioning of neural circuits and molecular pathways in the brain that support social behavior? How does social experience interact with information in the genome to modulate brain activity? Here, we address these questions by highlighting progress that has been made in identifying and understanding two key "vectors of influence" that link genes, the brain, and social behavior: (i) Social information alters gene expression in the brain to influence behavior, and (ii) genetic variation influences brain function and social behavior. We also discuss how evolutionary changes in genomic elements influence social behavior and outline prospects for a systems biology of social behavior.

Genes and social behavior have long had a tempestuous relationship in both science and society, and the "nature-versus-nurture" debate still has its adherents. This controversy persists because the relations among genes, the brain, and social behavior have complex entanglements across several different time scales (*1*), ranging from organismal development and physiology all the way to evolutionary time (Fig. 1). Genes do not specify behavior directly but rather encode molecular products that build and govern the functioning of the brain through which behavior is expressed. Brain development, brain activity, and behavior all depend on both inherited and environmental influences, and there is increasing appreciation that social information can alter brain gene expression and behavior. Furthermore, variation in behavior shapes the evolution of genomic elements that influence social behavior through the feedback of natural selection.

What is social behavior? Animals perform many activities during the course of their lives with the goal of surviving and reproducing: they find food and mates, defend themselves, and in many cases care for their offspring or other relatives. These activities become social when they involve interactions among members of the same species in a way that influences immediate or future behavior. One of the fundamentals of social behavior is communication (*2*). Diverse social behaviors involve the production, reception, and interpretation of signals that influence individual behavior in a manner that depends on social context.

Given the diversity and complexity of social behavior, is it realistic to anticipate that conserved mechanisms and general principles operate to control social behavior at the genomic level? We believe so. Although specific behavioral outcomes vary widely from species to species, the biological needs that drive these behaviors are deeply shared. Social behavior has clearly evolved multiple times, but probably within a framework of conserved neural mechanisms. All systems of social behavior share the following features: (i) They are acutely sensitive and responsive to social and environmental information. (ii) This information is transduced within individual organisms by one or more primary sensory pathways. (iii) The transduced neural signals are processed and integrated in specific circuits of the brain via conserved signal transduction and neuromodulatory systems. (iv) The resulting internal state of the animal ultimately controls its behavior.

Understanding the relations between genes and social behavior is especially challenging, because methods of experimental genetics have not been developed for animal species with the most compelling social repertoires—such as songbirds, cichlid fish, social insects (featured in this Review), and voles [discussed in Donaldson and Young's Review in this issue (*3*)]. Fortunately, through progress in whole-genome sequencing and comparative genomics, "model social" species are taking their place alongside the classic model genetic species in molecular analyses of behavior (*4*). It is now possible to compare model social species that vary in behavior and brain activity with one or more large-scale technologies (transcriptomics, epigenomics, proteomics, metabolomics, etc.). Results can be readily translated to model genetic species such as the mouse (*Mus musculus*) or the fruit fly (*Drosophila melanogaster*) for sophisticated genetic manipulations (*5*, *6*). Biologists no longer have to choose to study either a model genetic or a model social species; instead, they can choose both. This underscores the importance of actively pursuing research on diverse organisms that can capture the full richness and range of social behaviors.

In this article, we review selected findings that illustrate the relations between genes, the brain, and social behavior. As an organizational heuristic, we highlight two "vectors of influence" (Figs. 2 and 3): Vector 1 (Fig. 2) describes how social information leads to changes in brain gene expression, brain function, and social behavior, and Vector 2 (Fig. 3) describes how genetic variation between individuals leads to variation in social behavior.

Social Influences on Brain Gene Expression (Vector 1)

The genome was once thought to be a relatively passive blueprint guiding organismal development. Recent results show that genomes in fact remain highly responsive throughout life to a variety of stimuli associated with social behavior. Social information can lead to changes in the brain and behavior through effects on the genome (Fig. 2).

The first demonstrations of gene responses to social stimuli focused on a handful of immediate early genes (*7*), and one of these has proven especially useful. Referred to now as *egr1* (*8*), this transcription factor–encoding gene was discovered and named independently (*ngfi-a*, *zif-268*, *krox-24*, *tis8*, *zenk*) in different species. A specific link to social behavior was first suggested by studies in songbirds (*9*). Songbirds engage in rich social interactions that are mediated by their songs, which are learned vocal signals. The structure of songbird society varies by species, ranging from territorial to colonial, but in all cases, songbirds recognize and discriminate individual conspecifics according to their vocalizations. In the male zebra finch (*Taeniopygia guttata*), the singing of another male bird induces *egr1* expression in a specific subregion of the auditory forebrain devoted to hearing (*9*). Not simply an auditory response, *egr1* expression in this region is specifically linked to the social importance of the signal. Pure tones or white noise are ineffective stimuli. Moreover, the *egr1* response varies with recent familiarity to the particular song; previously unheard songs elicit strong responses, whereas familiar songs elicit little or no response (*10*). This has given rise to the speculation that the function of the genomic response is to help the brain update its natural representations in a changing social environment (*7*). A familiar song probably represents a familiar individual, whereas an unfamiliar song may represent a potentially threatening intruder.

The *egr1* response is also enhanced when the bird is listening in the presence of conspecifics, compared with when he is alone. This provides a neuromolecular analog of the "audience effect," a phenomenon in which an individual's performance depends on whether it is alone or with others (*11*). Other social interactions trigger *egr1* responses in other regions of the songbird brain, and the magnitude of the response can vary according to the intrinsic sociality of the species (*12*) and the immediate context of the experience (*13*).

egr1 was also the focus of another marked demonstration of gene responses in the brain resulting

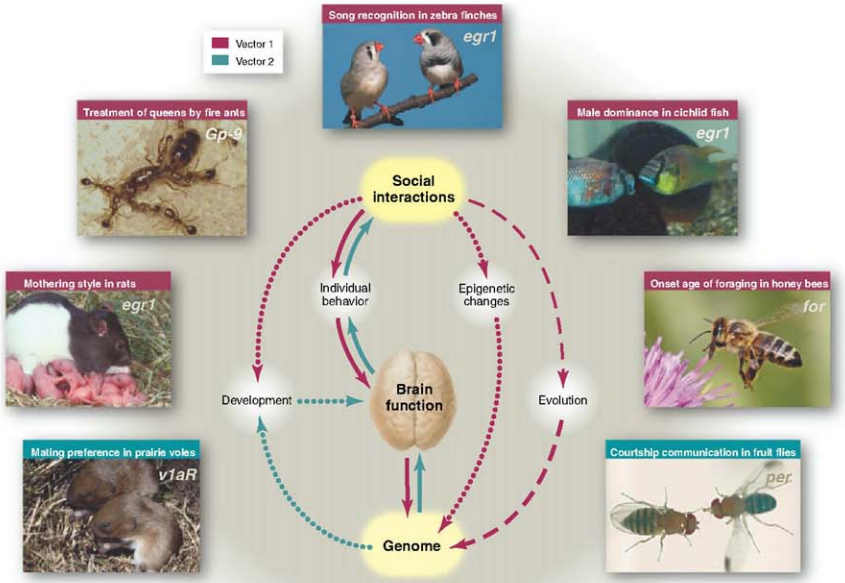
[1]Department of Entomology, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, IL 61801, USA. [2]Neuroscience Program, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, IL 61801, USA. [3]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Avenue, Urbana, IL 61801, USA. [4]Department of Biology and Neurosciences Institute, Stanford University, Stanford, CA 94305–5020, USA. [5]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, 601 South Goodwin Avenue, Urbana, IL 61801, USA.

*To whom correspondence should be addressed. E-mail: generobi@illinois.edu

from recognition of social opportunity in a highly social cichlid fish (*Astatotilapia burtoni*) (*14*). In many animal societies, dominance hierarchies structure all social interactions; position in the hierarchy governs access to resources that determine who reproduces and how often. *A. burtoni* has an elaborate dominance hierarchy, reinforced both by aggressive fighting and the ability of dominant males to ascertain relative rank by observation alone, using transitive inference to determine which male in a group is most dominant (*15*). Subordinate males have reduced fertility. When the alpha male is removed from a group, a subordinate male quickly starts to exhibit dominant behavior. In this social ascent, he displays dramatic changes in body coloration and behavior. Within minutes, but after the onset of behavioral change, *egr1* is induced specifically in the hypothalamic anterior preoptic area in neurons containing gonadotropin-releasing hormone (GnRH), a peptide critical for reproduction. These neurons increase in size and degree of dendritic arborization, and they also increase expression of GnRH mRNA and protein. These cellular and molecular responses depend on the recognition of a social opportunity and ascension to dominance; they are not elicited in individuals who are already dominant. Because *egr1* is a transcription factor, it is likely that these effects on the GnRH neurons are direct, but this has not yet been demonstrated. These results show that social information also can lead to changes in behavior that transiently alter patterns of brain gene expression (a variant of Vector 1).

Although *egr1* is only one of many socially responsive genes (see below), its molecular and cellular character provides insights of general importance. First, *egr1* can be induced by brief experiences, its expression reaching a peak 20 to 60 min later, in a "genomic action potential" (*7*). Second, *egr1* can immediately suppress or enhance the transcription of other genes, depending on which proteins it interacts with in different cell types (*16*). Third, results with *egr1* suggest how social experience might trigger changes in larger gene networks in the brain. By means of the application of high-throughput technologies for measuring the expression of many genes simultaneously, it is now clear that responses to social stimuli can be massive, involving hundreds or thousands of genes and perhaps many different brain regions at once.

In one of the first such studies, microarrays were used to measure brain gene expression patterns in the honey bee (*Apis mellifera*) at distinct life stages, finding expression differences in thousands of genes (*17*). Worker honey bees change jobs as they age. They spend the first 2 to 3 weeks of their adult life working in the hive caring for the brood, maintaining the nest and other activities, and then shift to collecting nectar and pollen outside the hive on behalf of their colony for the remainder of their 4- to



CREDIT: (FINCHES) P. WEGNER/PETER ARNOLD INC.; (FISH) RUSSELL FERNALD; (HONEY BEE) WIKIPEDIA; (FRUIT FLIES) C. T. ZHU, S. D. YEH, J. K. TRUE; (VOLES) GEORGE MCCARTHY/CORBIS; (RATS) M. KERN/PETER ARNOLD INC.; (FIRE ANTS) REN ROSS/UNIVERSITY OF GEORGIA

Fig. 1. Complex relationships connect genes, the brain, and social behavior. These relationships operate over three time scales: (i) physiological time via effects on brain activity (solid lines), (ii) developmental time via slower effects on brain development and genome modification (dotted lines), and (iii) evolutionary time via the processes of natural selection (dashed line). Arrow colors refer to Figs. 2 and 3 (pink, Fig. 2; blue, Fig. 3), which provide details about the nature of these interactions. Images depict some of the animals and genes featured in this review, clockwise from top: zebra finch (*T. guttata*), cichlid fish (*A. burtoni*), honey bee (*A. mellifera*), fruit fly (*D. melanogaster*), prairie vole (*M. ochrogaster*), rat (*R. norvegicus*), and fire ant (*S. invicta*). The genes listed (in italics on the photographs) are responsive to social interactions as described in the text.

6-week life. Despite this fixed pattern of behavioral maturation, the precise age when a bee redirects its energies from work in the hive to foraging depends on its perceptions of the colony's needs, which are communicated in part by pheromones. For example, if a bee colony loses a large fraction of its foraging force, some of the younger bees can speed up their rate of maturation and become "precocious" foragers. This occurs because inhibitory pheromones produced by older, foraging-age bees become less available. Although it looks like a bee foraging on a flower is a solitary affair, the onset age of foraging is subject to strong social influence. Perception of bee pheromones alters the expression of hundreds of genes in the bee brain over a period of days to weeks (*18*). Particularly affected are genes encoding transcription factors (*18*) and metabolic proteins (*19*).

Shifts in the expression of large populations of genes during social experience are also being observed in microarray-based experiments in fish and songbirds, and in these cases the alterations are both large and rapid. In the swordtail fish, *Xiphophorus nigrensis*, different social experiences quickly induce distinct patterns of gene response expression when measured at the level of the whole brain (*20*). For example, some genes are turned on in females as they interact with attractive males but are off when they interact with other females, and vice versa. In zebra finch song-recognition experiments, thousands of other RNAs (in addition to *egr1*) increase or decrease in the auditory forebrain within 30 min of the onset of an unfamiliar song stimulus (*21*). A day after the song has been entrained by repetition, however, the same now-familiar song no longer induces the behavioral response or the "unfamiliar" molecular response. But an altogether previously unknown and different gene expression profile has now emerged, which may represent the new baseline state.

These observations suggest that social information can have large global effects on gene expression in the brain, perhaps best described as shifts in neurogenomic states rather than as activation of particular genes in local neural circuits. A future challenge will be to confront the anatomical complexity of the brain, to describe and understand these genomic states at both finer and coarser scales of anatomy and time. For example, a single neuron may exist in different functional states as a result of modulation of synaptic proteins, which can alter the efficiency by which transient synaptic signals are consolidated into stable, lasting cellular changes (*7*). In neuronal circuits and ensembles, changes in the expression of ion channel proteins could affect how quickly the cell can respond and lead to changes in network function (*22*). At the whole-brain level, changes in the global neurogenomic state in the brain (such as the familiar daily sleep-waking cycle) involve massive changes in gene expression throughout the brain as a function of behavioral state (*23*).

Social Influences on Brain Gene Expression: Long-Term Epigenetic Effects

Social signals can also trigger long-lasting epigenetic modifications of the genome. These are heritable changes in expression of specific genes that are not attributable to changes in DNA sequence (Fig. 2). This phenomenon was first discovered in the transgenerational transmission of mothering style in rats (*Rattus norvegicus*) (*24*). Female rats that lick, groom, and nurse their pups extensively have offspring that are less responsive to stress and more responsive to their own pups. In contrast, pups that received less attention from their mothers are more easily stressed and show reduced responsiveness to their offspring.

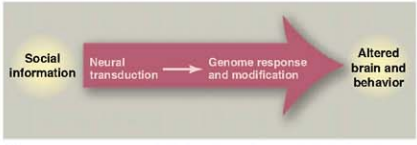


Fig. 2. Vector 1: From social information to changes in brain function and behavior. Social information is perceived by sensory systems and transduced into responses in the brain. Social information leads to developmental influences often mediated by parental care, as well as acute changes in gene expression that cause diverse effects (e.g., changes in metabolic states, synaptic connections, and transcriptional networks). Social information also can cause epigenetic modifications in the genome. Variation in both environment (V_E) and genotype (V_G) influences how social information is received and transduced and how these factors themselves interact ($V_E \times V_G$).

Because these differences in responsiveness to stress can be passed from generation to generation, they had previously been assumed to be inherited via traditional genetics. Instead, they stem from the fact that frequent mother/pup contact triggers at least two epigenetic changes in DNA methylation and very likely many more such molecular events. Methylation of the promoter region of the glucocorticoid receptor gene (which binds glucocorticoid stress hormones) allows the protein product of the *egr1* gene discussed above to up-regulate glucocorticoid receptor expression, especially in the hippocampus (*25*). In addition, methylation of the α1b promoter region of the estrogen receptor gene results in the up-regulation of estrogen receptors in the hypothalamus (*26*). Together with the results presented above for zebra finches, cichlid fish, and honey bees, these findings demonstrate that social experience can induce a range of changes in brain gene expression, from brief to enduring.

To date, epigenetic effects associated with social behavior have been studied at only a few genetic loci, but it is likely that many genes are similarly affected, especially in gene regulatory networks in the hypothalamic-pituitary-gonadal axis, known to be important for the regulation of a variety of vertebrate social behaviors (*12*, *27*). Genome-wide assays of epigenetic changes, assessing different regions and cell types in the brain, are necessary to fully understand how specific epigenetic modifications can both influence and be caused by social behavior.

Genotype-Environment Interactions and Social Behavior

The effects of social information on brain function and social behavior differ among individuals as a result of genetic variation between individuals (Fig. 2) (*28*). Such interactions between genotype and environment must always be accounted for in molecular analyses of social behavior. This has become an important focus in medicine, as psychiatric geneticists have been searching for genotype-environment interactions that might help to shed light on a wide range of psychiatric disorders. Some disorders, such as autism and depression, reflect social behavior gone awry. Evidence for genotype-environment interactions in psychiatric disorders has been reported, but these kinds of studies are still in their infancy (*29*–*31*).

Evidence of genotype-environment interactions influencing both social behavior and gene expression has been found in the fire ant *Solenopsis invicta* (*32*). Fire ants, like honey bees, live in colonies with thousands of workers, but while honey bee colonies have just a single queen, fire ant colonies can have one or more. The tendency to have either one or more queens has a genetic basis in fire ants. A genetic locus has been identified, *General Protein 9* (*Gp-9*), that is involved in regulating a key aspect of fire ant social organization—namely, the treatment of queens by the workers. Homozygous *BB* queens are larger and more fecund than *Bb* queens, and *BB* workers will only accept a single *BB* queen, resulting in one-queen colonies. *Bb* workers will accept multiple *Bb* queens, resulting in larger multiqueen colonies that are ecologically more invasive. *BB* workers become tolerant of multiple *Bb* queens when they are in colonies containing mostly *Bb* workers. In contrast, *Bb* workers do not change queen tolerance when they are in colonies containing mostly *BB* workers. *BB* workers in a *Bb* colony take on a *Bb* gene expression profile; comparable studies for *Bb* workers in *BB* colonies have not been done. For *BB* workers in *Bb* colonies, gene expression profiles are more strongly affected by colony genotype than their own genotype.

From Genes to Social Behavior (Vector 2)

Genetic or behavioral variants, either within populations of the same species or between species, offer an opportunity to understand how genetic

information influences the development and function of neural circuits and molecular pathways that mediate social behavior (Fig. 3). With the explosion in genome sequencing, the detection and analysis of genomic variation is becoming more routine, even in species without extensive histories of genetic analysis. However, such a comparative approach is particularly effective when direct experimental manipulation of genes or molecular pathways can be incorporated in the analysis to study behavioral consequences. A now-classic precedent is seen in the study of courtship communication in fruit flies [other aspects of fly courtship neurogenetics are reviewed in this issue by Dickson (*33*)]. Courtship involves some of the fundamentals of social behavior, even if it occurs fleetingly and in species with otherwise relatively solitary lifestyles, such as fruit flies. Like other more complex forms of social behavior, courtship involves communication between conspecifics to collect and process critical information—in this case about species, gender, receptivity, and quality of a potential mate.

Drosophila has a courtship song produced by the wings that is characterized by species-specific temporal coding. A comparison of two *Drosophila* species led to the identification of a specific difference in the *period* gene that was correlated with temporal differences in song structure. Transferring a small piece of the *period* gene from *D. melanogaster* to *D. simulans* caused the *melanogaster* males to produce the *simulans* call, rather than the *melanogaster* call (*34*). Thus, manipulation of even a single gene can have profound effects on behaviors associated with reproductive success.

Species differences were also exploited to study the molecular basis of mating preferences in the monogamous prairie vole, (*Microtus ochrogaster*) in comparison with the polygamous montane vole (*M. montanus*). As discussed elsewhere in this issue (*3*), sequence variation in the 5′ region of the vasopressin receptor gene *v1aR* causes differences, both in where this gene is expressed in the brain and in mating preferences. Recent findings showing no such relation between genetic variation and monogamy in other vole species (*35*) provide an excellent opportunity to explore how changes in different components of signaling pathways might result in similar changes in social behavior.

Behavioral variants within populations of *D. melanogaster* led to the discovery of another gene involved in social behavior in honey bees. Regulatory polymorphisms in *foraging* (*for*) are implicated in inter-individual genetic differences in foraging behavior, because flies with higher levels of *for* expression forage more actively than flies with lower levels of *for* expression (*6*). Foraging in *Drosophila* is not a social behavior, but these findings led to analysis of *for* orthologs in social insects. In social insects, differences in *for* expression are related to social activity, rather than genetic differences between individuals. In honey bees, brain *for* expression is higher in foragers than in hive bees, and socially induced precocious foragers show a precocious increase in *for* brain expression. Pharmacological treatment that activates the *for* pathway causes precocious foraging (*36*). Socially induced changes in *for* expression may be widespread in social insects (*37*).

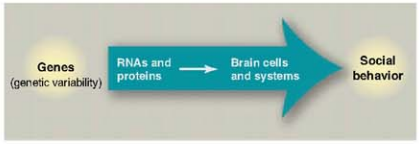


Fig. 3. Vector 2: From genes to social behavior. Genes influence the social behavior of an individual through their effects on brain development and physiology. This linkage is sensitive to both genetic (V_G) and environmental (V_E) variation and to their interactions ($V_G \times V_E$).

Much like *for* but in vertebrates, variation in the forkhead box P2 gene (*foxp2*) influences behaviors that have important social roles in multiple species, including human speech (*38*) and other forms of animal communication (*39*–*41*). *for* encodes guanosine 3′,5′-monophosphate–dependent protein kinase, and *foxp2* encodes a developmentally important transcription factor. Genes like *for* and *foxp2* may function as elements in a developmental or neural toolkit for building the circuits and systems underlying diverse socially embedded behaviors (*37*), even though they do not directly encode social behavior in any mechanistic sense.

For genes like *for* and *foxp2*, the link between gene and social behavior may be best appreciated by considering the evolutionary time scale (Fig. 1, pink dashed line). Through selection, genes may evolve according to their effects on a social behavior, even if their mechanistic roles in the neural expression of that behavior are subtle and indirect. Effects of selection may be detected in several aspects of comparative genome sequence data, including differences in amino acid codon frequencies, regulatory sequences, and gene copy number (*42*). Molecular evolution algorithms can be used to determine whether genes such as *for* and *foxp2* have been subjected to positive selection in particular lineages (*43*). Such analyses provide important tools for understanding the evolution of genes and other genomic elements that influence social behavior.

Prospects

Some progress has been made in understanding the specific relationships between genes and social behavior in a few species, but this enterprise is still in the formative stages. Understanding the molecular basis of social behavior presents a formidable intellectual challenge for several reasons. First, because behavior is diverse, assorted species must be used to extract the general principles that govern the molecular bases of social behavior. Dissecting a complex behavior into its components can help to identify root similarities across distantly related species (*4*). But even if deep molecular conservation is found among diverse species, one important issue remains. How can molecular pathways involved in behavior be conserved even when species show major differences in brain structure and the overall organization of the nervous system?

The second challenge is that the molecular path linking genes and behavior is invariably complicated (*44*). There are many levels of neural and neuroendocrine regulation that lie between the genome and a social behavior, including transcription, translation, posttranslational modifications, epigenetic changes, brain metabolism, neural (electrochemical) activity, and neuromodulation. Moreover, this regulation occurs in complex and dispersed temporal and spatial patterns within the brain, over physiological time, developmental time, and throughout an individual's life. The study of social behavior adds an additional tier of complexity because it depends on interactions and communication among individuals. In most cases, social behavior must be studied in a natural context in which the full repertoire of environmental influences and behaviors are expressed.

Despite the challenges, genetic and genomic approaches hold great promise for elucidating the molecular basis of social behavior. We have reasonably detailed knowledge of the two physical substrates responsible for behavior: the brain and the genome. We have a strong and growing arsenal of large-scale technologies and increasingly sophisticated methods of systems biology to profile changes in the brain during social behaviors. The time is ripe to combine this knowledge and these tools to aim for a comprehensive understanding of social behavior in molecular terms.

References and Notes

1. G. E. Robinson, *Science* **304**, 397 (2004).
2. D. Floreano, S. Mitri, S. Magnenat, L. Keller, *Curr. Biol.* **17**, 514 (2007).
3. Z. Donaldson, L. J. Young, *Science* **322**, 900 (2008).
4. G. E. Robinson, C. M. Grozinger, C. W. Whitfield, *Nat. Rev. Genet.* **6**, 257 (2005).
5. L. J. Young, R. Nilsen, K. G. Waymire, G. R. MacGregor, T. R. Insel, *Nature* **400**, 766 (1999).
6. M. B. Sokolowski, *Nat. Rev. Genet.* **2**, 879 (2001).
7. D. F. Clayton, *Neurobiol. Learn. Mem.* **74**, 185 (2000).
8. T. A. Eyre *et al.*, *Nucleic Acids Res.* **34**, D319 (2006).
9. C. V. Mello, D. S. Vicario, D. F. Clayton, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6818 (1992).
10. S. Dong, D. F. Clayton, *Genes Brain Behav.* **7**, 802 (2008).
11. C. Vignal, J. Andru, N. Mathevon, *Eur. J. Neurosci.* **22**, 949 (2005).
12. J. L. Goodson, *Horm. Behav.* **48**, 11 (2005).

13. E. D. Jarvis, C. Scharff, M. R. Grossman, J. A. Ramos, F. Nottebohm, *Neuron* **21**, 775 (1998).
14. S. S. Burmeister, E. D. Jarvis, R. D. Fernald, *PloS Biol.* **3**, e363 (2005).
15. L. Grosenick, T. S. Clement, R. D. Fernald, *Nature* **445**, 429 (2007).
16. K. J. O'Donovan, W. G. Tourtellotte, J. Millbrandt, J. M. Baraban, *Trends Neurosci.* **22**, 167 (1999).
17. C. W. Whitfield, A.-M. Cziko, G. E. Robinson, *Science* **302**, 296 (2003).
18. C. M. Grozinger, N. Sharabash, C. W. Whitfield, G. E. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14519 (2003).
19. S. A. Ament, M. Corona, H. S. Pollack, G. E. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4226 (2008).
20. M. E. Cummings *et al.*, *Proc. R. Soc. London Ser. B. Biol. Sci.* **275**, 393 (2008).
21. K. Replogle *et al.*, *BMC Genomics* **9**, 131 (2008).
22. H. Liu, M. M. Wu, H. H. Zakon, *Dev. Neurobiol.* **67**, 1289 (2007).
23. C. Cirelli, C. M. Gutierrez, G. Tononi, *Neuron* **41**, 35 (2004).
24. F. A. Champagne, D. D. Francis, A. Mar, M. J. Meaney, *Physiol. Behav.* **79**, 359 (2003).
25. I. C. Weaver *et al.*, *Nat. Neurosci.* **7**, 847 (2004).
26. D. L. Champagne *et al.*, *J. Neurosci.* **28**, 6037 (2008).
27. D. W. Pfaff, *Drive* (MIT Press, Cambridge, MA, 1999).
28. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Longmans Green, Harlow Essex, UK, ed. 4, 1996).
29. A. Caspi, T. E. Moffitt, *Nat. Rev. Neurosci.* **7**, 583 (2006).
30. D. K. Lahiri, B. Maloney, *Nat. Rev. Neurosci.* 10.1038/nrn2022-c1 (2006).
31. A. Abbott, *Nature* **454**, 154 (2008).
32. J. Wang, K. G. Ross, L. Keller, *PLoS Genet.* **4**, e1000127 (2008).
33. B. J. Dickson, *Science* **322**, 904 (2008).
34. D. A. Wheeler *et al.*, *Science* **251**, 1082 (1991).
35. S. Fink, L. Excoffier, G. Heckel, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10956 (2006).
36. Y. Ben-Shahar, A. Robichon, M. B. Sokolowski, G. E. Robinson, *Science* **296**, 741 (2002).
37. A. L. Toth, G. E. Robinson, *Trends Genet.* **23**, 334 (2007).
38. W. Enard *et al.*, *Nature* **418**, 869 (2002).
39. W. Shu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9643 (2005).
40. S. Haesler *et al.*, *PLoS Biol.* **5**, e321 (2007).
41. M. Groszer *et al.*, *Curr. Biol.* **18**, 354 (2008).
42. J. Sebat *et al.*, *Science* **316**, 445 (2007), published online 14 March 2007; 10.1126/science.1138659.
43. J. D. Jensen, A. Wong, C. F. Aquadro, *Trends Genet.* **23**, 568 (2007).
44. J. C. Hall, *J. Neurogenet.* **17**, 1 (2003).
45. We thank J. Desjardins, A. Fernald, K. A. Hughes, D. B. Kelley, K. Maruska, C. Olin, M. B. Sokolowski, L. J. Stubbs, members of the Clayton and Robinson laboratories, and two anonymous reviewers for reviews of this manuscript and C. Harrell for graphical assistance. Research by the authors cited here was supported by the following grants: NIH R01 NS051820 and NS045264 (D.F.C.); NIH NS34950 Javits Award and United States–Israel Binational Science Foundation 200596 (R.D.F.); and NIH R01 GM073644, NSF Frontiers in Biological Research EF04-25852, U.S. Department of Agriculture AG2003-35302-13490, and a Burroughs Wellcome Fund Innovation Award (G.E.R.).

10.1126/science.1159277

REVIEW

Oxytocin, Vasopressin, and the Neurogenetics of Sociality

Zoe R. Donaldson[1] and Larry J. Young[1,2]*

There is growing evidence that the neuropeptides oxytocin and vasopressin modulate complex social behavior and social cognition. These ancient neuropeptides display a marked conservation in gene structure and expression, yet diversity in the genetic regulation of their receptors seems to underlie natural variation in social behavior, both between and within species. Human studies are beginning to explore the roles of these neuropeptides in social cognition and behavior and suggest that variation in the genes encoding their receptors may contribute to variation in human social behavior by altering brain function. Understanding the neurobiology and neurogenetics of social cognition and behavior has important implications, both clinically and for society.

Social interactions affect every aspect of our lives, from wooing a mate and caring for our children to determining our success in the workplace. Abnormal manifestations of social behavior, such as the pathological trusting associated with Williams-Beuren Syndrome (*1*), social withdrawal in depression, and decreased social cognition in autism, profoundly affect the lives of those who suffer from these disorders. Neuroscientists once considered social behavior to be too hopelessly complex to understand at a mechanistic level, but advances in animal models of social cognition and bonding, as well as application of new technologies in human research have demonstrated that the molecular basis of social behavior is not beyond the realm of our understanding. There appears to be marked conservation in the molecular mechanisms regulating social behavior across diverse species, including our own.

Interacting with other neurotransmitter systems within specific neural circuits, neuropeptides have emerged as central players in the regulation of social cognition and behavior. Neuropeptides may act as neurotransmitters, if released within synapses, or as neurohormones, activating receptors distant from the site of release, which provides evolutionary flexibility to their actions (*2*). Within vertebrates, a majority of work relating neuropeptides to social behavior has focused on members of the oxytocin/vasopressin family. Homologs of oxytocin and vasopressin existed at least 700 million years ago and have been identified in such diverse organisms as hydra, worms, insects, and vertebrates. Among these distant taxa, oxytocin- and vasopressin-related peptides play a general role in the modulation of social and reproductive behaviors. In contrast to this apparent conservation in function, the specific behaviors affected by these neuropeptides are notably species-specific.

Only recently have scientists begun to dissect the roles of oxytocin, vasopressin, and their related receptors in human social behavior. Whereas human social behavior is more nuanced and complex than the behaviors typically assayed in other animals, this complexity has created unique opportunities to design finely honed tasks that have revealed a potential role for these peptides in personality, trust, altruism, social bonding, and our ability to infer the emotional state of others. Here, we review the evidence of evolutionary conservation within the vasopressin/oxytocin peptide family, briefly discuss the role of these peptides and their respective receptors in modulating social behavior and bonding, and provide a synthesis of recent advances implicating the oxytocin and vasopressin systems in human trust, cooperation, and social behavior.

Conservation of Neuropeptide Systems Regulating Social Behavior

The mammalian oxytocin and vasopressin nonapeptides, so called for their nine–amino acid composition, differ from each other at only two amino acid positions (Fig. 1). Oxytocin, vasopressin, and their respective nonmammalian vertebrate lineages are thought to have arisen from a gene-duplication event before vertebrate divergence. Within these lineages, peptides vary by a single amino acid, and their genes are found near each other on the same chromosome. Invertebrates, with few exceptions, have only one oxytocin/vasopressin homolog, whereas vertebrates have two (*3*, *4*).

In mammals, oxytocin and vasopressin are produced primarily within hypothalamic brain regions and then shuttled to the pituitary for peripheral release or projected to various brain regions. Notably, just as oxytocin and vasopressin are expressed within the hypothalamus of mammals, their homologs are expressed within similar neurosecretory brain regions of organisms as diverse as worms and fish. A characterization of annepressin (the homolog of oxytocin/vasopressin in segmented worms) and vasotocin (vasopressin's counterpart in bony fish) revealed conserved neural expression of these genes

[1]Center for Behavioral Neuroscience, Yerkes Regional Primate Research Center, Emory University, Atlanta, GA 30322, USA. [2]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA 30329, USA.

*To whom correspondence should be addressed. E-mail: lyoun03@emory.edu

within sensory-neurosecretory cells expressing common transcription factor combinations and tissue-restricted microRNAs (*5*). Furthering the idea that vasopressin/oxytocin homologs have ancient gene regulatory features that direct their expression in an evolutionarily conserved neural architecture, transgenic rats with a genomically integrated blowfish locus containing the isotocin gene, the teleost homolog of oxytocin, express isotocin within oxytocinergic neurons of the hypothalamus (*6*). These isotocin transgenic rats also show conserved physiological regulation of the transgene; osmotic challenge, a potent regulator of both oxytocin and vasopressin release, is sufficient to induce a sixfold increase in isotocin expression. This finding has been replicated in transgenic mice carrying the blowfish isotocin locus and the bovine oxytocin locus (*7*, *8*). These impressive results provide evidence that both the cis- and trans-acting elements controlling oxytocin/vasopressin neural expression, as well as its physiological regulation, are highly conserved across both vertebrates and invertebrates.

Oxytocin and vasopressin's roles in facilitating species-typical social and reproductive behaviors are as evolutionarily conserved as their structure and expression, although the specific behaviors that they regulate are quite diverse (*9*). For instance, conopressin, the snail homolog of oxytocin/vasopressin, modulates ejaculation in males and egg-laying in females. Within vertebrates, the distinct oxytocin and vasopressin peptide lineages often show sexually dimorphic expression and behavioral effects (*10*). The oxytocin lineage of peptides influences female sociosexual behaviors including sexual intercourse, parturition, lactation, maternal attachment, and pair bonding. Conversely, vasopressin typically influences male reproduction and behavior. Vasopressin is involved in erection and ejaculation in species including humans, rats, and rabbits (*11*, *12*), and it mediates a variety of male-typical social behaviors including aggression, territoriality, and pair bonding in various species. This sexual dichotomy in function is not universal, however, as it is becoming increasingly clear that both peptides have behavioral roles in males and females.

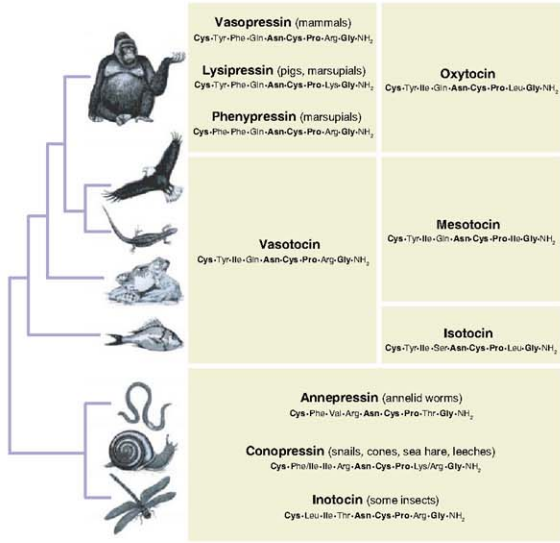


Fig. 1. Oxytocin and vasopressin homologs.

Oxytocin, Nurturing, and Social Attachment

The reproductive actions of oxytocin have been documented for over a century, and even in humans, studies identified the peripheral release of oxytocin during parturition, lactation, and sexual function as early as the 1950s. However, it was not until the 1970s and 1980s that scientists discovered the extent of oxytocin's involvement in behavior. In rats, central infusion of oxytocin stimulates maternal behavior in virgin females who would ordinarily ignore or attack pups. Conversely, experimental manipulations that decrease oxytocin levels or block oxytocin receptor activation within the brain reduce maternal behaviors (*3*).

In contrast to the induction of a generalized maternal state in rodents, maternal bonding in other species, including sheep and humans, consists of both a nurturing component and development of a selective attachment between the mother and her offspring. In sheep, oxytocin release in response to vaginocervical stimulation during parturition independently induces nurturing behaviors and facilitates the mother-infant bond selectivity after birth. Although oxytocin-induced maternal nurturing is mediated by some of the same brain regions in rodents and sheep, oxytocin also modulates maternal-infant bond selectivity in sheep by altering neurotransmitter activity within the olfactory bulb, essentially "priming" the olfactory systems for selective learning of offspring scent (*13*).

Humans display a great diversity of social attachments, one of which is selective preference for a particular mate, known as a pair bond. Pair bonding is exclusive to the 3 to 5% of mammalian species that are socially monogamous, and traditional laboratory organisms, such as rats and mice, do not display mate-based pair bonds. Instead, studies of monogamous prairie voles have yielded extensive insight into the molecular basis of pair bonding (*14*). Similar to its role in maternal bonding, central oxytocin administration facilitates partner-preference formation, a laboratory proxy of pair bonding, whereas blockade of the oxytocin receptor inhibits partner-preference formation in female prairie voles. This suggests that over evolutionary time within this species, a system specialized for maternal bonding in females was co-opted to modulate mate-directed bonds as well (*15*).

Selective bonding, including pair bonding and some mother-infant bonding, is hypothesized to result from concurrent neuropeptide modulation of pathways regulating reward and reinforcement and those involved in processing social information (*16*). Despite normal olfactory abilities, oxytocin knockout mice are unable to recognize previously encountered conspecifics, suggesting a specific role for this neuropeptide in the processing of social cues. In female prairie voles, blockade of either oxytocinergic or dopaminergic signaling within the reward- and reinforcement-associated nucleus accumbens prevents the development of a partner preference. Investigation of human pair bonding is still in its infancy, and there is no clear evidence that oxytocin contributes to pair-bond formation in humans.

Vasopressin and the Genetic Bases for Variation in Social Behavior

Even though both oxytocin and vasopressin show a conserved role in modulating social behavior in

general, the specific behaviors they influence show extensive variation among different species (*9*, *17*). For instance, vasopressin administration stimulates behaviors associated with monogamy, such as paternal care, mate guarding, and a selective preference for one's mate in male prairie voles; similar treatment does not induce these behaviors in nonmonogamous species. Likewise, in birds, the vasopressin homolog, vasotocin, increases vocalization and aggressive behavior in territorial male field sparrows but has only weak effects on aggression in colonial zebra finches. These species-specific behavioral effects are thought to be mediated, in part, by variation in brain receptor patterns rather than differences within the peptides (*18*).

Both oxytocin and vasopressin receptors display marked diversity in brain expression patterns. Oxytocin has a single identified receptor, whereas vasopressin acts in the brain on its two centrally expressed receptor subtypes, V1a and V1b. Of these two receptors, V1a plays a more prominent role in vasopressinergic modulation of social behavior and thus has been the focus of most research examining vasopressin's role in regulating social behavior (*15*, *19*). Male monogamous prairie voles, unlike polygamous meadow and montane voles, display selective mating-induced partner preferences, care for offspring, and selective aggression toward conspecifics. The development of these behaviors in male prairie voles is vasopressin-dependent. The brain distribution of vasopressin V1a receptor between these species is as divergent as their social behavior, and two experiments highlight the importance of these receptor patterns in mediating behavioral differences in these species (Fig. 2). First, simply increasing receptor expression within the reward and reinforcement circuitry in the meadow vole brain via viral vector-mediated gene expression enables individuals in this nonmonogamous species to form a selective preference for their mate, indicating that V1a receptor patterns influence a species' sociobehavioral repertoire (Fig. 2) (*20*). Along these same lines, transgenic insertion of the prairie vole V1a receptor gene and 2.7 kb of 5′ flanking sequence into the mouse genome leads to a prairie vole–like receptor pattern and altered social behavior (*21*). Together, this work highlights a potential evolutionary mechanism for creating behavioral diversity by altering receptor gene expression patterns. This idea is supported by investigation of individual differences in receptor patterns and behaviors within prairie voles. Individual voles, like humans, show differences in behaviors associated with monogamy, such as fidelity, space use, and paternal care. These behavioral differences are associated with pronounced variation in brain V1a receptor patterns, suggesting that receptor patterns modulate some aspects of both inter- and intraspecies behavioral diversity (*22*, *23*).

The genetic mechanisms underlying the phylogenetic and individual plasticity in V1a receptor expression in the brain and social behavior have begun to be explored. One potential candidate for generating diversity in V1a receptor gene (*avpr1a*) expression is a highly polymorphic, complex, repeat-containing DNA element known as a microsatellite, located in the 5′ flanking region of the *avpr1a* gene (*21*) (Fig. 3A). There are dramatic species differences and more subtle individual variation in this microsatellite element in voles, which are sufficient to drive in vitro expression differences in reporter gene assays (*24*). In vivo, when prairie voles with the longest and shortest microsatellite alleles are bred to homozygosity, the resulting progeny show brain V1a receptor expression and social behavior that differs according to *avpr1a* microsatellite length (Fig. 3A) (*22*). However, it remains unclear whether this region has effects on monogamous phenotypes in naturalistic settings (*25*). Although variation in *avpr1a* microsatellite length alone may not explain the evolution of the monogamous mating strategy in voles (*26*), these findings do suggest that polymorphisms in the *avpr1a* locus may contribute to both species differences and individual variation in V1a receptor distribution in the brain and consequently social behavior.

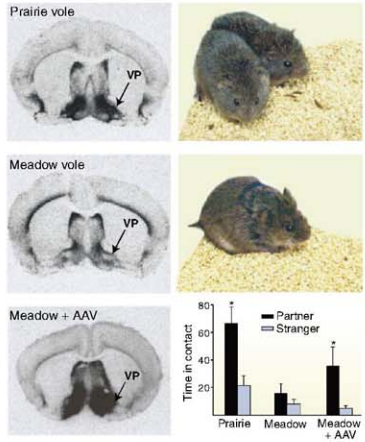


Fig. 2. Autoradiograms of vasopressin V1a receptor patterns in the ventral pallidum (VP) of socially monogamous prairie voles and polygamous meadow voles. When V1a receptor levels are artificially increased within the VP of meadow voles using adeno-associted viral vector (AAV) gene transfer (meadow + AAV), they display social behavior that is reminiscent of that of monogamous prairie voles, preferring social contact with their partner over a stranger (*20*). Error bars indicate SE; asterisks indicate $P < 0.05$. Time in contact is given in minutes.

Neurogenetics of Variation in Human Social Behavior

A number of recent findings suggest that variation in the *AVPR1A* locus may also contribute to sociobehavioral diversity in humans. Four polymorphic microsatellites, three within the 5′ flanking region and one within the intron of the gene, have been characterized and used in gene association studies. Various *AVPR1A* alleles have been directly associated with differences in human social behavior, personality traits relevant to social interaction, and the onset of reproduction (*27*, *28*). One study of 203 individuals has even found an association between the length of the most extensively studied of these polymorphisms, RS3, and altruism, a trait arguably necessary for successful formation of societies (*27*). Using an established economic game, researchers found that participants with longer V1a microsatellite alleles allocated more funds to another individual, despite the participant receiving as real money any unallocated funds at the end of the game.

Most recently, investigators asked the relevant question of whether *AVPR1A* genetic variability contributes to differences in human pair bonding among a cohort of 552 Swedish twin pairs, all of whom were living with a partner (*29*). Eighteen questions were used to probe partner bonding, perceived marital problems, and marital status. In particular, one allelic variant of a microsatellite in the 5′ flanking region of *AVPR1A*, allele RS3 334, was associated with significantly lower scores on the partner bonding scale in males only. Males who are homozygous for this allele were twice as likely to have experienced marital problems or threat of divorce and half as likely to be married if involved in a committed relationship. The presence of this allele in the male partner also correlated with perceived relationship quality in their female partner, suggesting the intriguing possibility that male *AVPR1A* genotype influences both partners.

In another line of research, *AVPR1A* variation has been hypothesized to specifically influence the sociobehavioral deficits characteristic of autism spectrum disorders. Three independent studies have identified associations between variants of this gene and autism. The most recent of these

studies more specifically identified an association between *AVPR1A* polymorphisms and socialization skills in high-functioning autistic participants in which language abilities were normal (*27*). It should be noted that these studies do not suggest that *AVPR1A* polymorphisms are a cause of autism, but they are consistent with the hypothesis that variation in this locus may be one contributor, among many others, to the social behavioral deficits associated with this spectrum of disorders. These studies should be viewed with caution, however, because the modest associations identified were with different alleles and haplotypes.

However, one allele is of particular interest: allele RS3 334. This allele, one of 16 different length variants in this repetitive region, was nominally implicated in autism in one study (*30*) and correlates most strongly with lower-quality partner bonding in males (*29*). A separate study also reported that individuals who carry this allele, as compared with all other alleles, have the highest levels of amygdala activation when performing an emotional face-matching task (*31*). When a different analysis on the same data was used, it revealed that longer microsatellites at this locus were associated with higher levels of amygdala activation during the face-matching task (Fig. 3B). Likewise, the only reported study examining *AVPR1A* expression in the brain in relation to polymorphisms has found that homozygous long RS3 microsatellite carriers have higher levels of V1a receptor hippocampal mRNA postmortem (*27*) (Fig. 3B).

A degree of skepticism should be maintained when considering any individual association study, and these studies require additional functional experiments examining the link between *AVPR1A* polymorphisms, gene expression, neural activity, and behavior. However, the repeated association of the *AVPR1A* locus with sociobehavioral traits and, in particular, the identification of allele RS3 334 in two independent studies has heightened interest in the hypothesis that variation in *AVPR1A* may contribute to variation in human sociobehavioral traits.

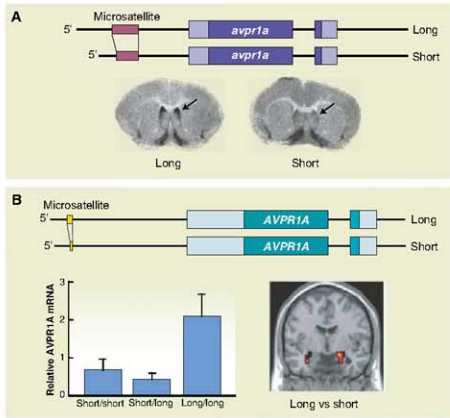


Fig. 3. Influence of genetic polymorphisms on gene expression levels, brain activation, and social behavior. **(A)** Within prairie voles, subtle microsatellite length variation upstream of the *avpr1a* transcription start site is associated with differences in V1a receptor expression patterns and behavior. **(B)** Allele length in an analogous microsatellite polymorphism upstream the human *AVPR1A* locus predicts V1a mRNA expression levels in the hippocampus (*27*), and longer alleles are also correlated with higher levels of amygdala activation during a face-viewing task (*31*). Error bars indicate SE.

Neuropeptides, Human Social Cognition, and Trust

Recent human studies have directly manipulated oxytocin and vasopressin systems by using intranasal administration to investigate the potential role for these peptides in modulating human social interactions. None of these studies have measured cerebrospinal levels of these peptides after intranasal infusion, but the reported behavioral effects of intranasal administration have been consistent, suggesting that whether acting peripherally or centrally, intranasally administered peptides do affect the brain and cognition.

Although the majority of these experiments have focused on oxytocin, limited studies with intranasal vasopressin have investigated its effects on social cognition. Human social inferences are derived largely from viewing facial expression, especially in the eye region. In human males, vasopressin administration decreases the perceived friendliness of faces and increases agonistic facial motor patterns (*32*). In contrast, females rate faces as friendlier and show affiliative facial motor programs after vasopressin application. Intranasal oxytocin has also been investigated in a similar paradigm, albeit in males only. When asked to categorize faces based on their expression, participants given intranasal oxytocin were better at classifying the emotions displayed on these faces and presumably inferring the mental state of another individual (*33*). Intranasal oxytocin infusion increases gaze to the eye region of human faces, providing a relatively simple potential mechanism for increasing the accuracy of mental state inference through increased information availability (*34*).

Complementary studies also support a role for oxytocin in modulating trust, thereby influencing cooperative interactions. Intranasal oxytocin increases the amount of money that an "investor" is willing to offer to a "trustee" who, after the amount is amplified by the experimenter, can then choose to return a smaller or larger sum back to the initial investor (*35*). Oxytocin does not, however, increase monetary allocations when the return on an investment is determined by a random lottery. This important control indicates that the effects of administration of this peptide are specific to the social interaction between the investor and trustee and therefore represents a quantifiable indication of interpersonal trust.

Two independent studies have now demonstrated the potential for maladaptive affects of oxytocin during social situations. In an extension of trust studies, researchers manipulated an investment scenario to simulate a situation in which the investor is "betrayed" by a trustee who returns less money than the initial investment (*36*). After the discovery of a betrayal of trust, the initial investment amounts decrease for placebo controls but not for oxytocin-treated investors. Similarly, in a different paradigm, pairing a shock with a face presentation skews the viewer's emotional rating of the face toward a more negative assessment, unless they have been administered oxytocin. In that case, they are likely to rate the shock-paired faces as more forgiving and sympathetic (*37*).

Insights into the neural mechanisms by which oxyotcin modulates social cognition have come from imaging studies that have consistently found that oxytocin decreases amygdala activity, regardless of the experimental scenario (*36*–*38*). The amygdala has been implicated in social information processing in both humans and animals, and bilateral amygdala lesions in humans impair their ability to judge the trustworthiness of others (*39*). As amygdala activation is also indicative of threatening or fearful stimuli, oxytocin mediated attenuation of amygdala activation may facilitate social interactions by decreasing potentially negative, anxiety-provoking associations.

Neuropeptides, Neurogenetics, and Society

Intranasal peptide administration and functional brain imaging studies are driving a revolution in our understanding of the roles of oxytocin and vasopressin in humans. However, our under-

standing is still extremely incomplete and hampered by both technical and design limitations. For instance, all oxytocin administration studies to date have been performed in males, and oxytocin's influence on bonding and social behavior in women has not been investigated. Furthermore, it is not known whether intranasal application of vasopressin or oxytocin mimics physiologically relevant events or represents pharmacological artifacts.

Among genetic studies, convergent evidence supports a role for the *AVPR1A* locus in modulating human social behavior, but the link between genes, the brain, and behavior remains weak. For instance, *AVPR1A* polymorphism is associated with differences in amygdala activation and autism, but its correlation with gene expression has only been investigated in the hippocampus. Finally, only one study has investigated the distribution of oxytocin and vasopressin receptors within the human postmortem brain (*40*), and techniques have improved since its publication. Development of selective positron emission tomography ligands for both oxytocin and vasopressin receptors will allow for in vivo studies of receptor expression and shed new light on correlations between genetic polymorphisms, brain receptor variability, and social cognition in humans. Although these limitations hinder our understanding of these neuropeptide systems, they are largely not insurmountable.

Many diseases, such as depression and social phobia, display symptomatic altered or deficient social behavior. In severe instances, such as autism and schizophrenia, abnormal social behavior is extremely debilitating. Identifying the molecular underpinnings of these social deficits may yield important clues into their treatment. For example, peripheral infusion of oxytocin increased retention of social cognition via enhanced emotional understanding of speech intonation and, unexpectedly, decreased repetitive behaviors (*41*). As peptides do not efficiently cross the blood/brain barrier, it is unclear how peripheral infusion of oxytocin mediates these effects, but these results are nevertheless intriguing. Even within healthy populations, social support enhances our ability to deal with stress and recover from disease. Oxytocin administration enhances the stress-alleviating effects of social support (*42*). The therapeutic potential of manipulating the oxytocin system remains to be explored in clinical trials, and the development of potent, selective agonists that penetrate the blood/brain barrier would be an important advancement toward this goal.

An understanding of the neurobiology of social behavior raises important questions for society. Should salesmen be allowed to use airborne oxytocin agonists to manipulate trust toward their own benefit? Should marital therapists include oxytocin infusions along with behavioral therapies to salvage relationships? Will genetic testing be used to screen potential partners or prospective sons-in-law? These and other questions may become the topics of discussion for bioethicists as we gain more insights into the neurobiology and neurogenetics of oxytocin, vasopressin, and sociality.

References and Notes

1. M. A. Martens, S. J. Wilson, D. C. Reutens, *J. Child Psychol. Psychiatry* **49**, 576 (2008).
2. M. Ludwig, G. Leng, *Nat. Rev. Neurosci.* **7**, 126 (2006).
3. H. K. Caldwell, W. S. Young III, in *Handbook on Neurochemistry and Molecular Neurobiology*, R. Lim, Ed. (Springer, New York, ed. 3, 2006), pp. 573–607.
4. R. Acher, J. Chauvet, M. T. Chauvet, *Adv. Exp. Med. Biol.* **395**, 615 (1995).
5. K. Tessmar-Raible *et al.*, *Cell* **129**, 1389 (2007).
6. B. Venkatesh, S. L. Si-Hoe, D. Murphy, S. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12462 (1997).
7. P. Gilligan, S. Brenner, B. Venkatesh, *J. Neuroendocrinol.* **15**, 1027 (2003).
8. M. Y. Ho, D. A. Carter, H. L. Ang, D. Murphy, *J. Biol. Chem.* **270**, 27199 (1995).
9. T. R. Insel, L. J. Young, *Curr. Opin. Neurobiol.* **10**, 784 (2000).
10. G. J. De Vries, G. C. Panzica, *Neuroscience* **138**, 947 (2006).
11. J. Gupta, R. J. Russell, C. P. Wayman, D. Hurley, V. M. Jackson, *Br. J. Pharmacol.* **155**, 118 (2008).
12. G. Segarra *et al.*, *J. Pharmacol. Exp. Ther.* **286**, 1315 (1998).
13. K. D. Broad, J. P. Curley, E. B. Keverne, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **361**, 2199 (2006).
14. C. S. Carter, A. C. DeVries, L. L. Getz, *Neurosci. Behav. Rev.* **19**, 303 (1995).
15. M. M. Lim, L. J. Young, *Horm. Behav.* **50**, 506 (2006).
16. L. J. Young, Z. Wang, *Nat. Neurosci.* **7**, 1048 (2004).
17. L. J. Young, *Horm. Behav.* **36**, 212 (1999).
18. J. L. Goodson, A. H. Bass, *Brain Res. Rev.* **35**, 246 (2001).
19. C. H. Hoyle, *Brain Res.* **848**, 1 (1999).
20. M. M. Lim *et al.*, *Nature* **429**, 754 (2004).
21. L. J. Young, R. Nilsen, K. G. Waymire, G. R. MacGregor, T. R. Insel, *Nature* **400**, 766 (1999).
22. E. A. D. Hammock, L. J. Young, *Science* **308**, 1630 (2005).
23. A. G. Ophir, J. O. Wolff, S. M. Phelps, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1249 (2008).
24. E. A. D. Hammock, L. J. Young, *Mol. Biol. Evol.* **21**, 1057 (2004).
25. A. G. Ophir, P. Campbell, K. Hanna, S. M. Phelps, *Horm. Behav.* 10.1016/j.yhbeh.2008.07.009 (2008).
26. S. Fink, L. Excoffier, G. Heckel, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10956 (2006).
27. S. Israel *et al.*, *Prog. Brain Res.* **170**, 435 (2008).
28. Z. M. Prichard, A. J. Mackinnon, A. F. Jorm, S. Easteal, *Hum. Mutat.* **28**, 1150 (2007).
29. H. Walum *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14153 (2008).
30. S. J. Kim *et al.*, *Mol. Psychiatry* **7**, 503 (2002).
31. A. Meyer-Lindenberg *et al.*, *Mol. Psychiatry* 10.1038/mp.2008.54 (2008).
32. R. R. Thompson, K. George, J. C. Walton, S. P. Orr, J. Benson, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7889 (2006).
33. G. Domes, M. Heinrichs, A. Michel, C. Berger, S. C. Herpertz, *Biol. Psychiatry* **61**, 731 (2007).
34. A. J. Guastella, P. B. Mitchell, M. R. Dadds, *Biol. Psychiatry* **63**, 3 (2008).
35. M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, E. Fehr, *Nature* **435**, 673 (2005).
36. T. Baumgartner, M. Heinrichs, A. Vonlanthen, U. Fischbacher, E. Fehr, *Neuron* **58**, 639 (2008).
37. P. Petrovic, R. Kalisch, T. Singer, R. J. Dolan, *J. Neurosci.* **28**, 6607 (2008).
38. P. Kirsch *et al.*, *J. Neurosci.* **25**, 11489 (2005).
39. R. Adolphs, *Nat. Rev. Neurosci.* **4**, 165 (2003).
40. F. Loup, E. Tribollet, M. Dubois-Dauphin, J. J. Dreifuss, *Brain Res.* **555**, 220 (1991).
41. J. A. Bartz, E. Hollander, *Prog. Brain Res.* **170**, 451 (2008).
42. M. Heinrichs, T. Baumgartner, C. Kirschbaum, U. Ehlert, *Biol. Psychiatry* **54**, 1389 (2003).

43. We thank A. Meyer-Lindenberg and R. Ebstein for help with assembling figures and L. McGraw, M. Modi, S. Subhabrata, H. Ross, and two anonymous reviewers for their comments on previous versions of the manuscript.

10.1126/science.1158668

REVIEW

Wired for Sex: The Neurobiology of *Drosophila* Mating Decisions

Barry J. Dickson

Decisions about whom to mate with can sometimes be difficult, but making the right choice is critical for an animal's reproductive success. The ubiquitous fruit fly, *Drosophila*, is clearly very good at making these decisions. Upon encountering another fly, a male may or may not choose to court. He estimates his chances of success primarily on the basis of pheromone signals and previous courtship experience. The female decides whether to accept or reject the male, depending on her perception of his pheromone and acoustic signals, as well as her own readiness to mate. This simple and genetically tractable system provides an excellent model to explore the neurobiology of decision making.

Behavior unfolds as animals select specific actions on the basis of sensory input, internal physiological states, and individual experience. A major goal for neuroscience is to understand how information processing and storage in neural circuits guides such action selection, and thus behavior. Genetic approaches in model organisms greatly facilitate the identification, characterization, and manipulation of individual circuit elements and can thereby establish causal relationships linking cellular biochemistry, circuit function, and animal behavior.

The sex life of the fruit fly *Drosophila melanogaster* is an ideal subject for such a study. Males decide whom to court, and females decide with whom to mate. The world-wide distribution and abundance of *Drosophila*, and its success as a

Research Institute of Molecular Pathology, Doktor Bohr-gasse 7, A-1030 Vienna, Austria. E-mail: dickson@imp.ac.at

genetic model organism, attest to the fly's exceptional ability to get these decisions right. Whatever the underlying neural mechanisms are, there is no doubt that they are robust and adaptive. They are also accessible to both genetic and physiological investigation at the level of single identifiable neurons.

Sturtevant first described the fly's mating behaviors almost a century ago (*1*). Benzer and, in particular, Hall have led the genetic investigation of these behaviors over the past few decades (*2*). Increasingly, the fly's sex life is now also attracting the attention of neurobiologists. Building on these behavioral and genetic studies, researchers are probing the anatomy and function of the neural circuits that guide the mating decisions of *Drosophila*. It is still early days, but work on the fly's mating decisions has the potential to reveal fundamental mechanisms of action selection—to teach us how the brain maps sensory input, internal states, and individual experience onto moment-to-moment behavioral choices.

The Male's Decision

Upon encountering another fly, a male must decide whether or not to court (Fig. 1). *Drosophila* males do not provide conspicuous nuptial gifts, and the courtship ritual itself may take only a few minutes. With such a modest investment, males are generally eager to try their luck. Nonetheless, there is a considerable reproductive benefit for males that can focus their efforts on those flies most likely to accept them: sexually mature female virgins of the same species. Evolution has endowed male flies with the innate ability to discriminate females from males and to court only females. Discriminating receptive from unreceptive females, however, is a skill acquired at least in part through trial-and-error learning.

Males rely primarily on chemical signals to detect suitable courtship objects, including both volatile pheromones detected by the olfactory system and nonvolatile pheromones detected by the gustatory system. If the male perceives pheromone signals predictive of mating success, he initiates an elaborate courtship ritual. A central component of this ritual is the courtship song produced by unilateral wing vibration. This song, or the visible wing extension that accompanies it, is an early and measurable readout of the male's decision to court. How then do pheromone signals, interpreted by a male brain, and in the context of previous courtship experience, guide the decision to sing?

Evaluating the evidence: Pheromone detection and processing. Many different pheromones have been shown to regulate *Drosophila* mating behaviors (*3*), but only for very few of these do we know the receptors and neurons that mediate pheromone detection. The best understood of these is the male pheromone *cis*-vaccenyl acetate (cVA), a volatile compound that modulates both male and female behavior. In males, detection of cVA suppresses courtship behavior, including the courtship song (*4*–*6*).

Flies detect odors through members of a large family of odorant receptors (ORs) that form heteromeric odor-gated ion channels (*7*–*10*). These receptors consist of a common Or83b subunit and a variable subunit that confers ligand specificity. The subunit that confers sensitivity to cVA is Or67d, expressed in a specific class of olfactory sensory neurons (OSNs) (Fig. 2A). Or67d is required in these neurons for cVA detection (*5*), and ectopic expression of Or67d in other OSNs renders them sensitive to cVA (*11*, *12*). Detection of cVA is facilitated by SNMP (sensory neuron membrane protein), a transmembrane protein of unknown function (*13*, *14*), and Lush, a secreted odorant binding protein (*15*). Lush binds cVA, and in doing so undergoes a conformational change (*16*). It is most likely this activated form of Lush, rather than cVA itself, that is the ligand for the Or67d:Or83b receptor.

OSNs of a specific class send axons to a discrete and stereotyped glomerulus in the antennal lobe, the insect analog of the mammalian olfactory bulb (*17*). These projections convert the peripheral map of odorant receptor activation into a spatial map of glomerular activation in the brain. This map is in turn conveyed to higher brain centers by the second-order olfactory projection neurons (PNs). Most of the OSN and PN circuitry has been mapped out at cellular resolution (*18*–*22*). The Or67d+ OSNs target a glomerulus called DA1 (*5*, *18*), where they faithfully pass the cVA signal on to the corresponding DA1 PNs (*23*, *24*) (Fig. 2A). The Or67d+ OSNs and DA1 PNs are both narrowly tuned to cVA, but the PNs are much more sensitive, presumably reflecting signal amplification due to the convergence of ~50 OSN inputs onto ~6 PNs. Both Or67d+ OSNs and DA1 PNs respond equally to cVA in males and females (*5*, *23*). However, the DA1 PNs form sex-specific arborizations in

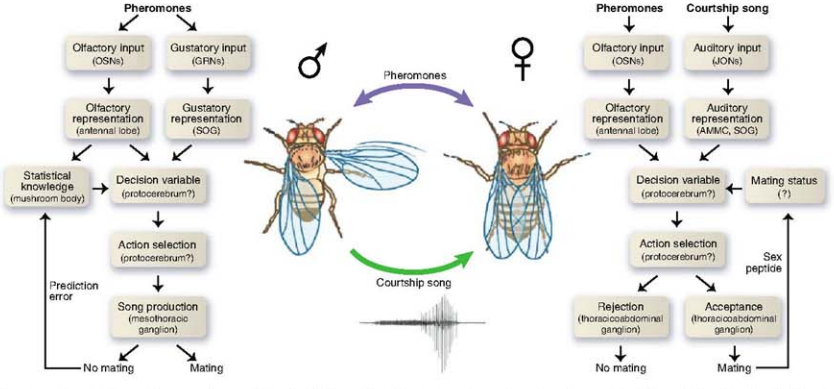


Fig. 1. Mating decisions. Elements of *Drosophila* male (**left**) and female (**right**) mating decisions. Parentheses indicate the relevant neurons or regions. OSNs, olfactory receptor neurons; GRNs, gustatory receptor neurons; JONs, Johnston's organ neurons; AMMC, antennal mechanosensory and motor center; SOG, suboesophageal ganglion. The decision variable reflects the likelihood of mating, constructed from sensory representations, acquired knowledge, and, for the female, current mating status. The decision variable guides a binary choice: for the male, to sing or not; for the female, to accept or reject the male.

the protocerebrum, hinting that they may feed the cVA signal into circuits that process it differently in males and females (*23*).

DA1 is one of three glomeruli that are larger in males than in females (*25*, *26*), analogous to the dimorphic glomeruli that receive pheromone input in the moth antennal lobe (*27*). The two other dimorphic glomeruli in *Drosophila* are VL2a and VA1v. The OSNs that innervate VL2a have not been identified, but VA1v has long been known as the target of OSNs that express the receptor subunit Or47b (*17*). Or47b confers sensitivity to fly odors produced by both sexes (*12*), and genetic perturbation of these neurons delays the onset of male courtship (*28*). This suggests that the Or47b/VA1v pathway may convey a pheromone signal that stimulates courtship toward either sex. This unidentified pheromone might, for example, be a species-specific stimulatory cue.

A simple model, then, is that sex and species might be encoded by the combined activity of the Or67d/DA1 and Or47b/VA1v pathways: *Drosophila melanogaster* female pheromones would activate the Or47b/VA1 pathway alone, whereas male pheromones would activate both pathways. Where in the brain would these signals be integrated? Current evidence suggests that pheromones, unlike fruit odors, are conveyed through the antennal lobe with little if any cross-channel processing (*24*). In the protocerebrum, however, the DA1 and VA1v PNs converge in a discrete region of the lateral horn that is spatially segregated from the region targeted by PNs that respond to fruit odors (*20*). This region is sexually dimorphic and may be the site at which the two pheromone signals are integrated to compute a sex-specific "decision variable" (*29*) that guides subsequent action selection (Fig. 1).

The male brain. Singing is a male-specific action. Females either do not select this action or cannot execute it. Such sex differences in neural function appear to be hard-wired during development (*30*). Sex in flies is primarily determined by the sex-specific splicing of two genes, *fruitless* (*fru*) and *doublesex* (*dsx*), both of which encode putative transcription factors (*31*–*33*). The expression and function of *fru*'s sex-specific transcripts is confined to the nervous system (*32*–*38*), whereas *dsx* acts in both neuronal and non-neuronal tissues (*39*). There is little overt dimorphism in the central nervous system, but numerous fine sex differences have been reported that depend on either *fru*, *dsx*, or both. A general rule may be that *dsx* controls neuroblast proliferation to produce initial differences in neuronal number (*40*, *41*), whereas *fru* acts in postmitotic neurons to regulate their survival or arborization patterns (*23*, *26*, *42*–*44*). The latter includes, for example, the dimorphic axonal arborizations of the DA1 PNs (*23*).

The sex differences sculpted by *fru*, but not those contributed by *dsx*, account for male-specific singing. Males that lack the male-specific fru^M isoforms do not sing (*33*, *38*), whereas those that lack dsx^M still do (*45*, *46*). Conversely, females forced to express fru^M sing (*47*), whereas those that express dsx^M do not (*45*, *48*). These observations justify the intense research focus on the set of neurons that express *fru* in efforts to understand how, in males, pheromone detection elicits singing.

There are ~2000 *fru*-expressing neurons in both sexes, including sensory, central, and motor neurons (*26*, *36*, *49*) (Fig. 2B). Among these are the OSNs and PNs that detect and process pheromones (*23*, *26*, *49*), motor neurons that regulate wing vibrations (*40*), and central neurons that contribute to the intervening neural processing (*43*, *50*). If the synaptic activity of all the *fru* neurons is blocked, all aspects of male courtship are suppressed, including song (*26*, *49*).

Classic studies of genetic sex mosaics (gynandromorphs) revealed that only a specific region in the dorsal protocerebrum must be genetically male for a fly to sing (*51*, *52*). The *fru* neurons in this region are therefore strong candidates to trigger courtship song. With modern genetic methods, this gynandromorph approach has now been extended to cellular resolution. If a small set of ~20 *fru* neurons in the dorsal brain, called the *fru*-P1 neurons, are masculinized in an otherwise normal female, then the female sings to other females (*43*). However, other neurons must also participate in the decision to sing, as many mosaic females without male *fru*-P1 neurons also sang in this study, and not all of those with male *fru*-P1 neurons did (*43*). Nonetheless, these findings suggest that the *fru*-P1 neurons are a critical element of the decision-making circuitry that triggers singing.

One implication of these gynandromorph studies is that the female thorax has the ability to sing, even though this action is never selected in a normal female. Indeed, direct optical stimulation of the thoracic *fru* neurons in headless flies ("flyPods") induces both males and females to sing (*50*). Evidently, a brief and artificial activation of thoracic *fru* neurons can kick-start local song-generating circuits present in both sexes. The same treatment in intact flies does not elicit robust singing, possibly because uniform activation of *fru* neurons in the brain generates conflicting inhibitory and stimulatory signals.

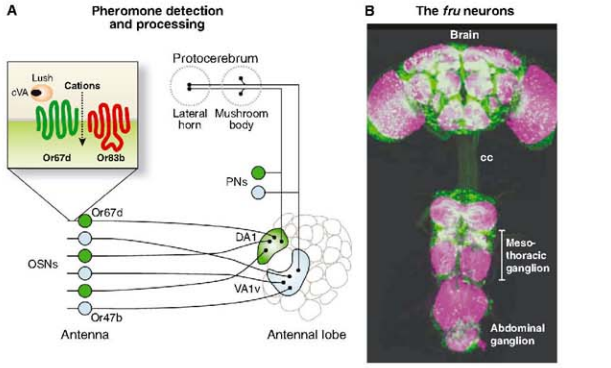


Fig. 2. Wired for sex. **(A)** cVA detection and processing in Or67d+ OSNs and DA1 PNs. **(B)** *fru*+ neurons in the CNS. Confocal image of the brain and ventral nerve cord of a fru^{GAL4} *UAS-PA-GFP* male, stained with nc82 to reveal synapses (magenta). Green fluorescent protein (GFP) fluorescence is shown in green. cc, cervical connective.

Normally, these signals might be generated in a sex-specific manner in the brain *fru* neurons, including perhaps the *fru*-P1 neurons.

The song produced by female flyPods is not a perfect rendition of the normal male courtship song. It is, however, significantly improved by expressing fru^M in all the *fru* neurons (*50*). This implies that fru^M contributes to the sexual differentiation of the circuits that produce the song as well as those that call them in to action. Correct male-specific differentiation of these song circuits also requires dsx^M, because the song is aberrant in flies that express fru^M but lack dsx^M, regardless of whether they are male (*46*) or female (*40*).

The emerging picture is that *fru* contributes to the sexual differentiation of neural circuits at all levels—sensory processing, action selection,

IMAGE PROVIDED BY ANNE VON PHILIPSBORN.

and action execution. Pronounced sex differences in behavior are due, however, to the differences that *fru* sculpts in circuits that select between alternative actions. In contrast, sensory and motor circuits may be largely common to both sexes but fine-tuned by *fru* and *dsx* for optimal performance according to the particular requirements of each sex. A consequence of this design is that perturbations that subvert or bypass the action selection circuits may readily cause one sex to inappropriately but poorly perform behaviors characteristic of the other.

Learning to predict mating success. Males and females differ reliably in their pheromone profiles, and evolution has encoded instructions for discriminating between them into the genome—instructions that are used to build hard-wired brain circuits. Obtaining a high rate of courtship success, though, also requires males to discriminate sexually mature virgins from immature females, unreceptive females that have recently mated, and females of other species. Because female pheromone profiles can vary substantially with time and place (*53*, *54*), the optimal classification scheme is not something that can easily be hard-wired into the brain. Innate mechanisms can implement a useful rule of thumb, but the optimal strategy for each location must be learned through experience. At least some circuit elements must remain plastic in order to record this experience. In this case, evolution has written into the genome the instructions for solving the classification problem, not the solution itself.

Evidence for this form of learning has come from experiments showing that males that experience courtship rejection by mated females are less inclined to subsequently court other mated females (*55*, *56*). Similarly, *Drosophila melanogaster* males experienced at courting *Drosophila simulans* females show suppressed courtship of other *simulans* females (*57*). In both cases, courtship of receptive *melanogaster* virgins is undiminished, indicating that the experienced male is indeed better able to discriminate receptive from nonreceptive females. This form of learning through trial-and-error interactions with the local environment is well modeled by classical reinforcement learning schemes (*58*).

The neural mechanisms that underlie this learning are still largely unknown. Dopamine neurons are strong candidates to convey reinforcement signals that report unexpected rejection and might be used to update circuits that compute or use the pheromone-based decision variable. The pheromone signals mediated by the Or67d/DA1 and Or47b/VA1v olfactory pathways are candidates for such experience-dependent modulation. The male pheromone cVA is transferred to females during mating (*59*, *60*) and could potentially be used to discriminate mated females from virgins. Indeed, cVA has been proposed to contribute to courtship learning, albeit not as a conditioned stimulus but rather as an unconditioned stimulus that suppresses the male's subsequent response to attractive female pheromones (*60*). This study, however, examined a general suppression of courtship toward all types of female that is sometimes observed when decapitated animals are used as test objects, not the selective suppression of courtship toward mated females that is observed with live animals (*55*, *56*, *61*, *62*).

A likely site for any experience-dependent modulation of pheromone responses is the mushroom body, a well-studied center for olfactory learning in insects (*63*). This protocerebral brain region receives input from both olfactory and reinforcement pathways. A specific class of mushroom body neurons—the γ neurons—express *fru*, and two separate lines of evidence have implicated these neurons in courtship learning. First, disrupting *fru* function in γ neurons blocks courtship suppression in short-term learning paradigms (*49*), although this has only been tested in assays for general rather than selective courtship suppression. Second, selective long-term courtship suppression requires the CPEB protein Orb2 specifically in γ neurons during or shortly after training (*61*). If mushroom body γ neurons are indeed the site for plasticity in pheromone processing, then we still need to identify the missing circuit elements that would integrate this signal with the lateral horn pathway so that past experience can guide future action.

The Female's Decision

Once the male decides to court, whether mating actually occurs is largely a matter of female choice (Fig. 1). The female decides on the basis of her assessment of her suitor's quality and on her own readiness to mate. If she decides to accept the male, she slows down and opens her vaginal plate for copulation. If not, she rejects the male by extruding her ovipositor in his direction, or simply flying away. How does the female select between these alternative actions, guided by stimuli from the male and her own internal state?

Assessing male quality. Female mating is in part stimulated by male pheromones, including cVA acting through the Or67d receptor (*5*). The most potent signal from the male, however, is his courtship song (*64*, *65*). Mute males (with clipped wings) have very little chance of courtship success (*1*, *66*), as do males that produce a poor song (*67*). The courtship success of mute males is, however, greatly improved by playback of a prerecorded song from a high-quality male (*66*, *68*). The song alone even induces lone females to slow down their movement (*69*, *70*), just as receptive females normally do in the presence of a singing male. The critical component of the courtship song is a series of short pulses, typically spaced about 35 ms apart (*68*, *71*, *72*). This interpulse interval is species-specific (*72*) and is a key factor in species recognition (*73*, *74*).

Flies detect sound through rotational movements of the antenna that are induced by the vibration of air particles (*75*). These movements activate exquisitely sensitive stretch receptor neurons in Johnston's organ, the fly's "ear," capable of detecting displacements of just a few nanometers (*76*). Just as the distinct classes of OSN in the olfactory system are each specialized to detect specific kinds of odorant, distinct classes of Johnston's organ neurons (JONs) may respond to distinct mechanical stimuli. Some may detect courtship song, but others may respond to vibrational signals important for other forms of acoustic communication (*77*), for flight control (*78*), or in gravity-sensing (*79*). In this regard, it is interesting to note that distinct subsets of JONs project to distinct regions in the brain (*80*), hinting at a functional segregation of mechanical signals analogous to the segregation of food odors and pheromones in the olfactory pathways. Many JONs are *fru*+ (*26*, *49*) and may have specific roles in the detection or processing of courtship song analogous to the specialization of *fru*+ OSNs for pheromone detection.

Female receptivity. How the female responds to these male signals depends on her own readiness to mate. Young virgins do not mate, nor do females that have recently mated. Adult females reach sexual maturity only when they are 1 to 2 days old (*81*). Immature virgins are, however, still attractive to males, who court them vigorously. This experience may provide females with an opportunity to learn about the quality of local males (*82*), just as males may use this experience to learn about local female pheromone profiles (*55*).

Once they mate, females store sperm for extended periods and use it efficiently (*83*). Unless they encounter a second male of substantially higher quality, there is little to be gained by mating again. It is also in the first male's interest that a female, once inseminated, does not readily mate again. This common interest has led to the evolution of a mechanism that renders females unreceptive after an initial mating, typically lasting until her sperm supply is depleted (*81*). Male seminal fluid contains a small peptide, the sex peptide (SP), that binds tightly to sperm (*84*, *85*). SP induces a suite of postmating responses in the female, including her reluctance to mate again. Females that mate to mutant males lacking SP readily mate again (*86*, *87*). Conversely, direct injection of SP into virgin females renders them unreceptive (*84*).

In the female, SP activates a specific receptor, SPR, a G protein-coupled receptor that activates the adenosine 3′,5′-monophosphate signaling pathway (*88*). Females lacking SPR remain receptive even after an initial mating or SP injection. SPR is broadly expressed in the nervous system, but its function is both necessary and sufficient within the *fru* neurons (*88*). Evidently, SP regulates female receptivity by modulating the properties of some subset of the *fru* neurons.

The female brain. Circuits somewhere in the female brain must integrate sensory inputs from the olfactory system, auditory system, and reproductive tract to decide between the alternative actions of accepting or rejecting the male (Fig. 1). These circuits remain largely uncharted. Classic gynandromorph studies have mapped a region of the dorsal brain that must be genetically female for a mosaic animal to be receptive (*89*), but modern methods for creating genetic mosaics have not yet been exploited to extend this analysis to cellular resolution.

It is possible that *fru* neurons in the brain are involved in female mating decisions, just as they are in male mating decisions. Silencing synaptic transmission of the *fru* neurons in females inhibits female receptivity (*90*), as does masculinizing them with fru^M (*47*). Similarly, although fru^F males (which produce female rather than male *fru* transcripts) are not particularly attractive to other males, they do show features of female behavior when they are occasionally courted (*91*). These fru^F males also behave like females when tested for aggression (*92*). Collectively, these findings strongly implicate at least some of the *fru* neurons in female behavior. However, none of these effects has yet been mapped to a specific subset of the *fru* neurons, and so it is still too early to tell whether they reflect the perturbed function of *fru* neurons in sensory, central, or motor processing. *fru* neurons clearly are involved in the sensory input relevant for female decision making, including *fru*+ pheromone-sensing neurons. The critical question, however, is whether *fru* neurons also contribute to the female decision-making circuits, and if so, whether these are the same as or different from those circuits that make mating decisions in the male. The classical gynandromorph studies suggest that they might be distinct (*89*).

Perspective

I have presented here our current understanding of the neural mechanisms that guide the mating decisions of male and female fruit flies. This picture contains many obvious gaps, and an urgent goal is to trace out the relevant neural circuits more completely and at cellular resolution. With the powerful genetic methods now available for circuit dissection (*93*, *94*), this should not take too long. Indeed, the precision with which individual circuit elements can be identified and manipulated is advancing much more rapidly than the methods for analyzing circuit function. Explaining behavior may be the ultimate goal, but behavior itself is a noisy and distal readout of circuit function. The behavioral output of genetically perturbed neural circuits will not always be meaningful, but the physiological properties of specific circuit elements may well be changed in highly informative ways.

This work should gradually reveal how chemical and auditory cues are detected and processed in the fly's brain, how these signals are interpreted in the context of internal physiological states and past experience, and how this information is used to make decisions that are fundamental to the animal's reproductive success. Will this teach us anything about information processing and storage in even more complex brains? Mating behaviors and their key regulatory genes evolve rapidly, and we cannot expect to extend these findings across vast evolutionary distances by homology. Neural networks may, however, be built by assembling simple and common modules into complex neuronal architectures (*95*, *96*). Similar architectures may be used to solve similar computational problems, even if the molecular mechanisms differ. The mammalian and insect olfactory systems, for example, use molecularly distinct receptor families to detect odors, and these odors trigger very different sets of behaviors. Yet, surprisingly, they process olfactory information in very similar ways (*97*). There may be only a limited set of efficient neural solutions to complex behavioral problems, including difficult decisions such as choosing a mate. Studying this process in the fly holds the promise of revealing how the computations performed by defined neural circuits can guide decision making and behavior, and how these computations emerge from the biochemical properties of the constituent neurons and their connections.

References and Notes

1. A. H. Sturtevant, *J. Anim. Behav.* **5**, 351 (1915).
2. J. Weiner, *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior* (Vintage Books USA, New York, 1999).
3. J. F. Ferveur, *Behav. Genet.* **35**, 279 (2005).
4. J. M. Jallon, C. Antony, O. Benamar, *C. R. Acad. Sci. Paris* **292**, 1147 (1981).
5. A. Kurtovic, A. Widmer, B. J. Dickson, *Nature* **446**, 542 (2007).
6. S. D. Mane, L. Tompkins, R. C. Richmond, *Science* **222**, 419 (1983).
7. R. Benton, S. Sachse, S. W. Michnick, L. B. Vosshall, *PLoS Biol.* **4**, e20 (2006).
8. E. M. Neuhaus *et al.*, *Nat. Neurosci.* **8**, 15 (2005).
9. K. Sato *et al.*, *Nature* **452**, 1002 (2008).
10. D. Wicher *et al.*, *Nature* **452**, 1007 (2008).
11. T. S. Ha, D. P. Smith, *J. Neurosci.* **26**, 8727 (2006).
12. W. van der Goes van Naters, J. R. Carlson, *Curr. Biol.* **17**, 606 (2007).
13. R. Benton, K. S. Vannice, L. B. Vosshall, *Nature* **450**, 289 (2007).
14. X. Jin, T. S. Ha, D. P. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10996 (2008).
15. P. Xu, R. Atkinson, D. N. Jones, D. P. Smith, *Neuron* **45**, 193 (2005).
16. J. D. Laughlin, T. S. Ha, D. N. Jones, D. P. Smith, *Cell* **133**, 1255 (2008).
17. L. B. Vosshall, A. M. Wong, R. Axel, *Cell* **102**, 147 (2000).
18. A. Couto, M. Alenius, B. J. Dickson, *Curr. Biol.* **15**, 1535 (2005).
19. E. Fishilevich, L. B. Vosshall, *Curr. Biol.* **15**, 1548 (2005).
20. G. S. Jefferis *et al.*, *Cell* **128**, 1187 (2007).
21. E. C. Marin, G. S. Jefferis, T. Komiyama, H. Zhu, L. Luo, *Cell* **109**, 243 (2002).
22. A. M. Wong, J. W. Wang, R. Axel, *Cell* **109**, 229 (2002).
23. S. R. Datta *et al.*, *Nature* **452**, 473 (2008).
24. M. L. Schlief, R. I. Wilson, *Nat. Neurosci.* **10**, 623 (2007).
25. Y. Kondoh, K. Y. Kaneshiro, K. Kimura, D. Yamamoto, *Proc. R. Soc. London B Biol. Sci.* **270**, 1005 (2003).
26. P. Stockinger, D. Kvitsiani, S. Rotkopf, L. Tirian, B. J. Dickson, *Cell* **121**, 795 (2005).
27. B. S. Hansson, H. Ljungberg, E. Hallberg, C. Lofstedt, *Science* **256**, 1313 (1992).
28. C. M. Root *et al.*, *Neuron* **59**, 311 (2008).
29. J. L. Gold, M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535 (2007).
30. B. I. Arthur Jr., J. M. Jallon, B. Caflisch, Y. Choffat, R. Nothiger, *Curr. Biol.* **8**, 1187 (1998).
31. S. E. Erdman, K. C. Burtis, *EMBO J.* **12**, 527 (1993).
32. H. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9687 (1996).
33. L. C. Ryner *et al.*, *Cell* **87**, 1079 (1996).
34. A. Anand *et al.*, *Genetics* **158**, 1569 (2001).
35. S. F. Goodwin *et al.*, *Genetics* **154**, 725 (2000).
36. G. Lee *et al.*, *J. Neurobiol.* **43**, 404 (2000).
37. G. Lee, A. Villella, B. J. Taylor, J. C. Hall, *J. Neurobiol.* **47**, 121 (2001).
38. A. Villella *et al.*, *Genetics* **147**, 1107 (1997).
39. K. C. Burtis, B. S. Baker, *Cell* **56**, 997 (1989).
40. E. J. Rideout, J. C. Billeter, S. F. Goodwin, *Curr. Biol.* **17**, 1473 (2007).
41. B. J. Taylor, J. W. Truman, *Development* **114**, 625 (1992).
42. J. C. Billeter *et al.*, *Curr. Biol.* **16**, 1063 (2006).
43. K. Kimura, T. Hachiya, M. Koganezawa, T. Tazawa, D. Yamamoto, *Neuron* **59**, 759 (2008).
44. K. Kimura, M. Ote, T. Tazawa, D. Yamamoto, *Nature* **438**, 229 (2005).
45. B. J. Taylor, A. Villella, L. C. Ryner, B. S. Baker, J. C. Hall, *Dev. Genet.* **15**, 275 (1994).
46. A. Villella, J. C. Hall, *Genetics* **143**, 331 (1996).
47. E. Demir, B. J. Dickson, *Cell* **121**, 785 (2005).
48. C. P. Kyriacou, J. C. Hall, *Science* **232**, 494 (1986).
49. D. S. Manoli *et al.*, *Nature* **436**, 395 (2005).
50. J. D. Clyne, G. Miesenbock, *Cell* **133**, 354 (2008).
51. J. C. Hall, *Behav. Genet.* **7**, 291 (1977).
52. F. von Schilcher, J. C. Hall, *J. Comp. Physiol. A* **129**, 85 (1979).
53. J. M. Jallon, J. R. David, *Evolution* **41**, 294 (1987).
54. J. F. Ferveur, M. Cobb, H. Boukella, J. M. Jallon, *Genetica* **97**, 73 (1996).
55. R. Dukas, *Anim. Behav.* **69**, 1203 (2005).
56. M. Reif, K. E. Linsenmair, M. Heisenberg, *Anim. Behav.* **63**, 143 (2002).
57. R. Dukas, *Behav. Ecol.* **15**, 695 (2004).
58. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 1998).
59. F. M. Butterworth, *Science* **163**, 1356 (1969).
60. A. Ejima *et al.*, *Curr. Biol.* **17**, 599 (2007).
61. K. Keleman, S. Kruttner, M. Alenius, B. J. Dickson, *Nat. Neurosci.* **10**, 1587 (2007).
62. S. M. McBride *et al.*, *Neuron* **24**, 967 (1999).
63. M. Heisenberg, *Nat. Rev. Neurosci.* **4**, 266 (2003).
64. T. A. Markow, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6200 (1987).
65. F. Rybak, G. Sureau, T. Aubin, *Proc. Biol. Sci.* **269**, 695 (2002).
66. F. von Schilcher, *Anim. Behav.* **24**, 622 (1976).
67. F. von Schilcher, *Behav. Biol.* **17**, 187 (1976).
68. H. C. Bennet-Clark, A. W. Ewing, *Nature* **215**, 669 (1967).
69. F. von Schilcher, *Anim. Behav.* **24**, 18 (1976).
70. S. A. Crossley, H. C. Bennet-Clark, H. T. Evert, *Anim. Behav.* **50**, 827 (1995).
71. H. H. Shorey, *Science* **137**, 677 (1962).
72. A. W. Ewing, H. C. Bennet-Clark, *Behaviour* **31**, 288 (1968).
73. H. C. Bennet-Clark, A. W. Ewing, *Anim. Behav.* **17**, 755 (1969).
74. C. P. Kyriacou, J. C. Hall, *Anim. Behav.* **30**, 794 (1982).
75. M. C. Göpfert, D. Robert, *J. Exp. Biol.* **205**, 1199 (2002).
76. J. T. Albert, B. Nadrowski, M. C. Gopfert, *Curr. Biol.* **17**, 1000 (2007).
77. M. Paillette, H. Ikeda, J. M. Jallon, *Bioacoustics* **3**, 247 (1991).
78. M. Gewecke, P. Schlegel, *J. Comp. Physiol. [A]* **67**, 325 (1970).
79. D. A. Baker, K. M. Beckingham, J. D. Armstrong, *J. Comp. Neurol.* **501**, 756 (2007).
80. A. Kamikouchi, T. Shimada, K. Ito, *J. Comp. Neurol.* **499**, 317 (2006).
81. A. Manning, *Anim. Behav.* **15**, 239 (1967).
82. R. Dukas, *Behav. Ecol.* **16**, 800 (2005).
83. M. C. Bloch Qazi, Y. Heifetz, M. F. Wolfner, *Dev. Biol.* **256**, 195 (2003).

84. P. S. Chen *et al.*, *Cell* **54**, 291 (1988).
85. J. Peng *et al.*, *Curr. Biol.* **15**, 207 (2005).
86. T. Chapman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9923 (2003).
87. H. Liu, E. Kubli, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9929 (2003).
88. N. Yapici, Y. J. Kim, C. Ribeiro, B. J. Dickson, *Nature* **451**, 33 (2008).
89. L. Tompkins, J. C. Hall, *Genetics* **103**, 179 (1983).
90. D. Kvitsiani, B. J. Dickson, *Curr. Biol.* **16**, R355 (2006).
91. R. Stoop, B. I. Arthur, *Chaos* **18**, 023123 (2008).
92. E. Vrontou, S. P. Nilsen, E. Demir, E. A. Kravitz, B. J. Dickson, *Nat. Neurosci.* **9**, 1469 (2006).
93. L. Luo, E. M. Callaway, K. Svoboda, *Neuron* **57**, 634 (2008).
94. B. D. Pfeiffer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9715 (2008).
95. R. Milo *et al.*, *Science* **298**, 824 (2002).
96. M. Reigl, U. Alon, D. B. Chklovskii, *BMC Biol.* **2**, 25 (2004).
97. R. I. Wilson, Z. F. Mainen, *Annu. Rev. Neurosci.* **29**, 163 (2006).
98. I thank R. Wilson, K. Feng, J. Yu, A. von Philipsborn, and M. Häsemeyer for comments on the manuscript.

10.1126/science.1159276

PERSPECTIVE

Searching for Genes Underlying Behavior: Lessons from Circadian Rhythms

Joseph S. Takahashi,[1,2,3]* Kazuhiro Shimomura,[2,3] Vivek Kumar[1,2,3]

The success of forward genetic (from phenotype to gene) approaches to uncover genes that drive the molecular mechanism of circadian clocks and control circadian behavior has been unprecedented. Links among genes, cells, neural circuits, and circadian behavior have been uncovered in the *Drosophila* and mammalian systems, demonstrating the feasibility of finding single genes that have major effects on behavior. Why was this approach so successful in the elucidation of circadian rhythms? This article explores the answers to this question and describes how the methods used successfully for identifying the molecular basis of circadian rhythms can be applied to other behaviors such as anxiety, addiction, and learning and memory.

In the 1970s, Seymour Benzer and his colleagues uncovered a remarkable number of genes that underlie neural and behavioral functions. They treated the fruit fly *Drosophila* with mutagens and systematically screened them for behavioral abnormalities (*1*, *2*). The discovery, in one of these screens, of flies with mutations in the *period* gene—which show longer or shorter cycles of the flies' endogenous 24-hour clock—by Konopka and Benzer (*3*) remains the exemplar for genetic dissection of behavior (*4*). Why was the search for circadian mutants so successful, and why were unbiased approaches to gene discovery so important?

Today, we understand the molecular mechanism of the circadian clock in a number of model organisms ranging from bacteria to humans (*5*, *6*). In retrospect, it is clear that the genes regulating circadian rhythms would not have been easily uncovered without the use of forward genetic screens (*2*). In each model organism (*Drosophila*, *Neurospora*, cyanobacteria, *Arabidopsis*, and mouse), previously unknown pathways were identified by the cloning of circadian mutants (*5*). Even today, with the benefit of complete genome sequences, the function of most of these "clock genes" would have been difficult to work out without those screens because our preconceived notions of the properties of a clock gene were largely incorrect. For example, a long history of anatomical and physiological experiments in mammals [beginning with the localization of the central clock, the suprachiasmatic nucleus (SCN), in 1972 (*7*)] indicated that clock genes should be tissue-specific and restricted to the SCN. In addition, it was assumed that clock genes would be transcribed in a circadian pattern. Both of these assumptions were incorrect, at least in part. The *Clock* and *Period* genes are expressed ubiquitously, and *Period*, but not *Clock*, is expressed in a circadian pattern (*8*–*10*). We now realize that clock genes are really housekeeping genes and are integral to the most basic functions of cells, and that virtually all cells in the body contain cell-autonomous circadian oscillators (*11*–*13*).

The genetic screens for circadian mutants were successful because we have a deep understanding of circadian phenotypes and robust assays. Of the measurable parameters of circadian rhythms (period, phase, and amplitude), the choice of circadian period as a primary phenotype has proven to be key. Circadian period length is a fundamental aspect of the clock that can be easily and accurately measured by 24/7 automated monitoring (*14*). Parameters such as amplitude are inherently ambiguous because they can be influenced by processes downstream of the circadian oscillatory system (that is, by output pathways). Phase measures can also be ambiguous because phase can be influenced by changes in input pathways that entrain the oscillator. Under steady-state conditions, period length (even when measured at the behavioral level) is directly correlated with the period of the underlying circadian pacemaker system (*15*) and thus is a very sensitive measure of the rate-limiting molecular steps in the circadian pathway.

Another key to success has been the accuracy with which circadian period length can be measured. The onset of rhythms in activity—in particular, wheel-running by rodents—is a remarkably precise phenotype (*14*). The inbred mouse strain C57BL/6J, for example, shows an average period length for circadian wheel-running in constant darkness of 23.7 hours with a standard deviation (SD) of 0.17 hours, or 10 min (*16*). This is a relative standard deviation (RSD = SD/mean) of only 0.72% (Fig. 1A and table S1). [The RSD of circadian rhythms of individual mice is even lower, about 0.2%, which is second in precision only to the neural oscillator driving the electric organs in fish (*17*).] Thus, in genetic screens, more than 99% of C57BL/6J mice have circadian periods between 23.2 and 24.2 hours (which represents ±3 SD from the mean); and any mutagenized mouse with a period outside this range is likely to be a mutant. (The precision of circadian rhythms in mice is strain-dependent, and C57BL/6J is one of the most precise for period length.) Indeed, phenotype-driven genetic screens based on period length have been the most successful for the discovery and functional assessment of circadian clock genes (*2*).

How can one apply what we have learned from circadian clock genetics to discover genes underlying other complex behaviors in the mouse? Over the past decade, my colleagues and I have systematically applied forward genetic screens in the mouse using the point mutagen *N*-ethyl-*N*-nitrosourea (ENU) to find mutants that affect learning and memory, anxiety, locomotion, vision, and response to psychostimulant drugs (*18*). To select appropriate screens, we looked for behaviors in which the phenotype was well established and for which we had an understanding of the neural loci and circuitry underlying the behavior. We required that the behavioral assay be amenable to automated data acquisition. The phenotypic screen also had to be scalable to achieve a throughput of more than 200 mice screened per week, so that ~10,000 mice could

[1]Howard Hughes Medical Institute, Northwestern University, Evanston, IL 60208, USA. [2]Department of Neurobiology and Physiology, Northwestern University, Evanston, IL 60208, USA. [3]Center for Functional Genomics, Northwestern University, Evanston, IL 60208, USA.

*To whom correspondence should be addressed. E-mail: j-takahashi@northwestern.edu

be screened per year, which is needed to have a reasonable probability of recovering mutants. This requirement for high throughput places restrictions on the types of behavioral assays one can employ. Tests such as the Morris water maze for learning and memory are too time- and resource-intensive (*19*). Thus, contextual and cue-dependent fear conditioning assays (in which animals learn to associate fear with a particular environment or signal) have been the most practical to use in assessing learning and memory (*18*). Although rodents will self-administer drugs, mimicking human addiction (*20*), this method is also too cumbersome to use for screening. Instead, locomotor responses induced by psychostimulants have been used for drug addiction screens (*18*).

The behavioral assays chosen for mutant screening should be valid indicators for the behavioral phenotype of interest. Contextual fear conditioning requires both a hippocampal pathway for memory and an amygdala pathway for the fear/emotional component (*21*), but the memory aspect of this task can be assessed separately. In our screens of mutagenized mice for genes required for long-term memory, we have tested for defects in the long-term component of fear conditioning (*22*). We already know that many long-term memory mutants in *Drosophila* are sensitive to gene dosage, indicating that these mutations may point to a rate-limiting step in the underlying memory process (*23*).

On the other hand, psychostimulant-induced locomotion and sensitization of this response are thought to be elements of the addictive response, but, unlike drug self-administration tests, this assay alone cannot assess addictive behavior (*20*). Thus, in such screens, one may or may not be able to recover mutants that affect addiction because addictive behavior is not directly measured.

The validity of the available screens may also be questionable for other behavioral phenotypes. For example, two common tests for anxiety, open-field behavior and the elevated-plus maze, use a measure of exploration or activity that can be easily influenced by many factors (such as ataxia or sensory changes) in addition to the anxiety phenotype that the tests are intended to assess. Thus, an important lesson learned from circadian mutants is that a deep understanding of the phenotype and a phenotypic measure (such as period length) that reflects rate-limiting processes are critical for the success of behavioral screens.

A second lesson, already well known from classical genetics studies in model organisms and humans and reinforced by the isolation of circadian mutants, is that point mutations can be more informative than loss-of-function mutations. Although loss-of-function mutations are necessary to prove that a specific gene is necessary for a phenotype, in practice such muta-

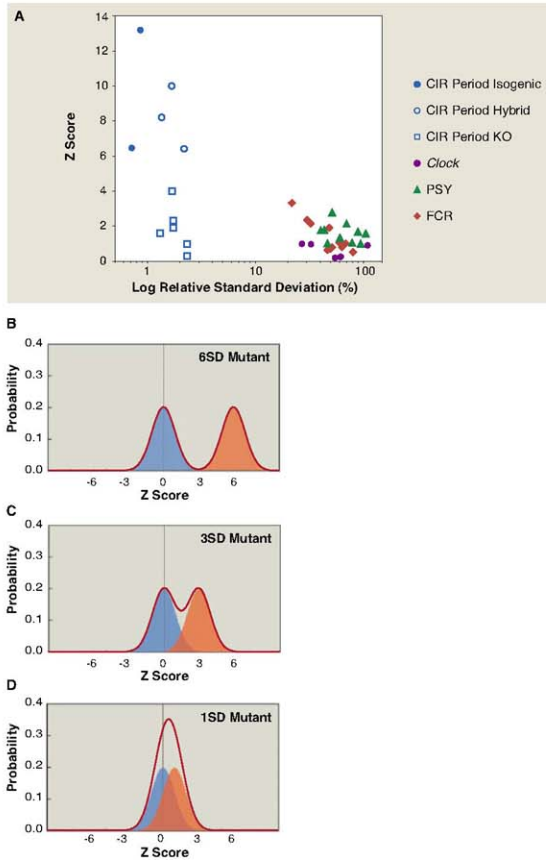


Fig. 1. (A) The relationship between the precision of the phenotypic assay and the phenotypic effect size (*Z*-score deviation from the wild type) for circadian rhythm, fear conditioning, and psychostimulant response mutants. The circadian period measurements (circles and squares) have the lowest RSD as compared to fear conditioning (FCR, diamonds) and psychostimulant response (PSY, triangles) mutants as well as other circadian measurements such as activity level and circadian amplitude (purple circles). Circadian point mutants (solid circles) discovered in isogenic forward genetic screens have the highest mutant *Z* scores and the lowest RSD values. The same mutations in hybrid genetic backgrounds have slightly higher RSDs. Circadian knockout (KO) mutations (open squares, hybrid genetic background) have lower mutant *Z* scores as compared to point mutants isolated in forward genetic screens. See table S1 for detailed information and references on mutants. **(B** to **D)** Phenotypic distribution of *Z* scores of wild-type (blue shading) and mutant (orange shading) animals that differ by 6, 3, or 1 SD, respectively, from the mean of the wild type. The graphs illustrate the expected 1:1 distribution of wild-type and mutant populations for a backcross, assuming a dominant mutation. The solid red line indicates the sum of the wild-type and mutant populations.

tions can be limiting because the complete loss of the gene is often lethal. In addition, paralogous genes may take over the function of the deleted gene, and there may be other compensatory mechanisms. Hypomorphic and dominant-negative mutations can circumvent some of these problems, and the recovery of an allelic series of mutations can also be informative for understanding the structure and function of genes. Indeed, the short-period, long-period, and loss-of-rhythm mutant alleles in the original *Drosophila period* locus indicated that *per* was not only essential for circadian behavior but also influenced the rate of the process (*3*). The strongest period-altering mutant alleles in mammals are caused by point mutations rather than loss-of-function alleles; for example, the *casein kinase 1e tau* (*Csnk1e*^{tau}) and the *Clock*$^{\Delta 19}$ mutations (*8*, *24*), which produced the highest mutant phenotypic deviations from the wild-type mean on a scale normalized to the SD (Z scores) when compared with those of null mutations (Fig. 1A). Loss-of-function mutations of *Clock* and *Csnk1e* have subtle phenotypes because paralogous genes have overlapping functions (*25*, *26*). Both *Csnk1e*^{tau} and *Clock*$^{\Delta 19}$ harbor dominant-negative mutations (*26*, *27*), and the dominant-negative action of these mutant alleles on paralogous genes explains their strong phenotypes. Given that there are worldwide efforts to create null mutations for every gene in the mouse genome in embryonic stem cells (*28*, *29*), point mutagenesis with ENU provides a complementary and mutually reinforcing approach for expanding the allelic spectrum of available mutations in the mouse (*29*–*31*).

A third lesson from circadian screens is that the variance of the behavioral phenotype is critical for the discovery of strong mutants. When we compare the precision of a phenotype using the relative standard deviation, the circadian period phenotype is almost two orders of magnitude more precise than assays for fear conditioning, psychostimulant response, or other circadian parameters such as amplitude or activity level (Fig. 1A). The precision of the assay was key for the isolation of circadian mutants such as *Clock*$^{\Delta 19}$ or *Overtime* (*Fbxl3*^{Ovtm}) that cause mutant period phenotypes that are 6 to 10 SD (*Z* scores of 6 to 10) away from those of wild-type mice (*16*, *32*) (Fig. 1A), where there is virtually no overlap in the period lengths of the mutant and wild-type populations. For this to be the case, the difference in the average values of the mutant and wild-type populations must exceed 6 SD (so that ±3 SD for each population will not overlap) (Fig. 1B). We have not recovered such strong mutations for other behavioral phenotypes in the mouse. The strongest effect sizes for fear conditioning mutants in mice are about 1 to 3 SD (*Z* scores of 1 to 3) (*33*) (Fig. 1A and table S1). Under these conditions, mutant and wild-type mice have significant phenotypic overlap (Fig. 1, C and D), and it is not possible to score individual mice as either mutant or wild type (unless each mouse is test-crossed, which is very costly). These phenotypic effect sizes are similar to those of strong quantitative trait loci (QTLs), which originate from naturally occurring allelic variants (*34*, *35*). In order to map such subtle mutants, quantitative genetic mapping approaches must be used because the mice cannot be scored qualitatively. The resolution of such crosses will be limited by the inherent ambiguity in quantitative versus qualitative scoring of phenotypic traits.

An additional problem is that mutations can be suppressed or modified by genetic background (*36*, *37*), a common issue with behavioral mutants. In order to locate ENU-induced mutations, the mice need to be crossed with other strains of mice that differ enough from the mutant strain to provide informative genetic markers. QTLs that are independent of the behavioral mutation can segregate in these mapping crosses, confounding mapping of the mutant locus (*35*, *38*). Thus, there is a tradeoff between choosing a strain for mapping that is divergent enough to provide sufficient polymorphic markers and the increased likelihood of QTLs arising from the divergent genetic background. Typically, one chooses other standard inbred mouse strains for mapping crosses because of the availability and high resolution of genetic markers. For example, in a typical cross of C57BL/6J to other common inbred strains such as A/J, BALBc/J, or DBA/2J, half of the available microsatellite or single-nucleotide polymorphism (SNP) markers would be polymorphic. The variance of the phenotype in these hybrid crosses will increase because QTLs from the parental strains (Haldane's rule) will segregate in crosses with the induced mutation, exacerbating the phenotypic overlap and confusing identification of the mutant mice. To circumvent some of these issues, particularly the increase in phenotypic variance and the suppression of mutant phenotypes, one can use very closely related mouse strains for mapping. This approach was not feasible until recently because of the paucity of genetic markers between very closely related strains; however, with the availability of high-density SNPs, even such related strains (C57BL/10J for mapping mutants isolated on a C57BL/6J isogenic background, for example) will have informative markers. The use of closely related strains can succeed in genetic mapping experiments when conventional approaches have failed.

A final consideration arises: If the phenotypic effect sizes of ENU-induced behavioral mutants are as subtle as QTLs, then why bother with induced mutants instead of attempting to clone QTLs for the same phenotype (*39*)? The answer to this question is simple. QTLs or naturally occurring allelic variants are extremely difficult to clone because there are so many sequence variants between the parental strains of mice [about one polymorphism every 440 base pairs (bp) in high-SNP regions] (*34*, *40*). Thus, a typical QTL mapping experiment will narrow the region responsible for a phenotypic variation to ~40 Mb (*34*, *35*), within which there will be literally tens of thousands of sequence variants. By contrast, with ENU-induced mutations isolated on an isogenic C57BL/6J strain background, there will be only about one sequence variant in 100,000 bp, and for overt (causative) mutations the rate is 10 times lower (one in a million bp of coding sequence) (*41*). The subtle phenotypic effect size (~1 SD), however, means that the mapping resolution of an ENU-induced allele and a QTL will be comparable because both must be treated as quantitative traits.

How can the mutation be found within an ~40-Mb region? Until recently, with a region this large, it would be extremely difficult to locate and clone the sequence alteration responsible for the behavioral defect unless one could find obvious candidate genes to analyze. However, massively parallel sequencing (*42*) and targeted selection of genomic sequences (*43*, *44*) now allow sequencing of all of the genes (exons and flanking regulatory regions) in a 50-Mb interval (400 to 450 genes on average in the mouse). The exome of 400 genes can be sequenced with current Roche 454 FLX sequencing methods at >25 times coverage in a single run; eight such samples could be sequenced in a single run by Illumina/Solexa (or Applied Biosystems SOLiD) sequencing at an even lower cost per sample. The sequence capacity of these instruments is already increasing, so that the cost of brute-force sequencing of the exome of all genes in a large genomic interval will be reasonable and will permit the positional cloning of essentially any ENU-induced mutation in the future.

The prospects for forward genetic and positional cloning approaches to complex behavior in the immediate future are bright. ENU-induced mutants isolated on isogenic mouse strains with available finished genomic sequence will enable the identification of critical behavioral genes by brute-force sequencing of all candidate genes in a large genomic interval with sequence-capture methods coupled with massively parallel sequencing. With these methods, even the most subtle behavioral mutants in the mouse can be successful targets for gene discovery, and the black box that underlies the genetic architecture of many complex behaviors will soon be opened.

References and Notes

1. S. Benzer, *Sci. Am.* **229**, 24 (1973).
2. J. S. Takahashi, *J. Biol. Rhythms* **19**, 339 (2004).
3. R. J. Konopka, S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2112 (1971).
4. J. S. Takahashi, L. H. Pinto, M. H. Vitaterna, *Science* **264**, 1724 (1994).
5. D. Bell-Pedersen *et al.*, *Nat. Rev. Genet.* **6**, 544 (2005).
6. J. S. Takahashi, H. K. Hong, C. H. Ko, E. L. McDearmon, *Nat. Rev. Genet.* **9**, 764 (2008).

7. D. R. Weaver, *J. Biol. Rhythms* **13**, 100 (1998).
8. D. P. King *et al.*, *Cell* **89**, 641 (1997).
9. Z. S. Sun *et al.*, *Cell* **90**, 1003 (1997).
10. H. Tei *et al.*, *Nature* **389**, 512 (1997).
11. E. Nagoshi *et al.*, *Cell* **119**, 693 (2004).
12. D. K. Welsh, S. H. Yoo, A. C. Liu, J. S. Takahashi, S. A. Kay, *Curr. Biol.* **14**, 2289 (2004).
13. S. H. Yoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5339 (2004).
14. S. M. Siepka, J. S. Takahashi, *Methods Enzymol.* **393**, 230 (2005).
15. C. S. Pittendrigh, S. Daan, *J. Comp. Physiol. A* **106**, 223 (1976).
16. M. H. Vitaterna *et al.*, *Science* **264**, 719 (1994).
17. K. T. Moortgat, T. H. Bullock, T. J. Sejnowski, *J. Neurophysiol.* **83**, 971 (2000).
18. M. H. Vitaterna, L. H. Pinto, J. S. Takahashi, *Trends Neurosci.* **29**, 233 (2006).
19. D. M. Bannerman, M. A. Good, S. P. Butcher, M. Ramsay, R. G. Morris, *Nature* **378**, 182 (1995).
20. M. Thomsen, S. B. Caine, *Behav. Genet.* **37**, 101 (2007).
21. E. A. Phelps, J. E. LeDoux, *Neuron* **48**, 175 (2005).
22. E. R. Kandel, *Science* **294**, 1030 (2001).
23. R. L. Davis, *Physiol. Rev.* **76**, 299 (1996).
24. P. L. Lowrey *et al.*, *Science* **288**, 483 (2000).
25. J. P. DeBruyne, D. R. Weaver, S. M. Reppert, *Nat. Neurosci.* **10**, 543 (2007).
26. Q. J. Meng *et al.*, *Neuron* **58**, 78 (2008).
27. D. P. King *et al.*, *Genetics* **146**, 1049 (1997).
28. The International Mouse Knockout Consortium, *Cell* **128**, 9 (2007).
29. Y. Gondo, *Nat. Rev. Genet.* **9**, 803 (2008).
30. A. Acevedo-Arozena *et al.*, *Annu. Rev. Genomics Hum. Genet.* **9**, 49 (2008).
31. B. T. Kile, D. J. Hilton, *Nat. Rev. Genet.* **6**, 557 (2005).
32. S. M. Siepka *et al.*, *Cell* **129**, 1011 (2007).
33. A. Matynia *et al.*, *PLoS One* **3**, e2121 (2008).
34. J. Flint, W. Valdar, S. Shifman, R. Mott, *Nat. Rev. Genet.* **6**, 271 (2005).
35. L. L. Peters *et al.*, *Nat. Rev. Genet.* **8**, 58 (2007).
36. W. F. Dietrich *et al.*, *Cell* **75**, 631 (1993).
37. K. E. Strunk, V. Amann, D. W. Threadgill, *Genetics* **167**, 1821 (2004).
38. K. Shimomura *et al.*, *Genome Res.* **11**, 959 (2001).
39. J. H. Nadeau, W. N. Frankel, *Nat. Genet.* **25**, 381 (2000).
40. K. A. Frazer *et al.*, *Nature* **448**, 1050 (2007).
41. D. R. Beier, *Mamm. Genome* **11**, 594 (2000).
42. E. R. Mardis, *Annu. Rev. Genomics Hum. Genet.* **9**, 387 (2008).
43. T. J. Albert *et al.*, *Nat. Methods* **4**, 903 (2007).
44. E. Hodges *et al.*, *Nat. Genet.* **39**, 1522 (2007).

45. Supported by NIH grants U01 MH61915, R01 MH078024, and P50 MH074924 (Silvio O. Conte Center) to J.S.T., Takeda Research grant 07-030R to K.S., and NIH National Research Service Award F32 DA024556 to V.K. J.S.T. is an Investigator in the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/909/DC1
Table S1

10.1126/science.1158822

PERSPECTIVE

Biology, Politics, and the Emerging Science of Human Nature

James H. Fowler* and Darren Schreiber

In the past 50 years, biologists have learned a tremendous amount about human brain function and its genetic basis. At the same time, political scientists have been intensively studying the effect of the social and institutional environment on mass political attitudes and behaviors. However, these separate fields of inquiry are subject to inherent limitations that may only be resolved through collaboration across disciplines. We describe recent advances and argue that biologists and political scientists must work together to advance a new science of human nature.

Aristotle is credited with being the first political scientist. In his work *The Politics* he carefully describes the constitutions of a number of different city-states, starting a science of political institutions that would last thousands of years. But he is also known for first asserting the biological uniqueness of human political behavior with his famous observation: "Man is, by nature, a political animal" (*1*).

It has not been easy for us to follow in his footsteps. In the past 50 years, biologists have learned a tremendous amount about human biology and its genetic basis. At the same time, political scientists have been intensively studying the effect of the social and institutional environment on political attitudes and behaviors. However, biologists and political scientists have been working largely in isolation of one another. Little cross-disciplinary work has been done.

This must change for two important reasons. First, recent evidence is making it increasingly clear that genetic variation plays an important role in explaining variation in human political behavior. Second, additional evidence in neuroscience indicates that the human brain may be adapted particularly to solve social problems that are explicitly political. Much of this evidence is associational, and we therefore should be cautious in using it to build causal theories. However, if the need for sophisticated social cognition drove the evolution of the human brain (*2*), then a new science of human nature will require comprehending human biology in a sociopolitical context.

Department of Political Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.

*To whom correspondence should be addressed. E-mail: jhfowler@ucsd.edu

Genes and Politics

Since at least the middle of last century, theories about political attitudes and behavior have focused almost exclusively on information about peer and parental socialization, socioeconomic factors, and political institutions. Although political scientists have made progress on important questions, their models have become burdened with dozens of ad hoc theories, and they usually fit poorly to the data (*3*). For example, one prominent model of voter participation includes 32 variables but accounts for only 31% of the variance in turnout behavior (*4*). Moreover, the theories underlying these empirical models typically ignore genetic or biological factors that might be responsible for the remaining variation.

Unbeknownst to most political scientists, psychologists and behavioral geneticists began using twin studies in the 1980s to study variation in social attitudes, and these studies suggested that both genes and environment played a role (*5*). However, this early work did not specifically pursue the question of whether political orientations were heritable, and political scientists remained largely unaware of the heritability of social attitudes until 2005. In that year, the *American Political Science Review* published a reanalysis of political questions on a social attitude survey of twins that suggested that liberal and conservative ideologies are heritable (*6*). Follow-up studies showed that genes did not play a role in the choice of a political party (*6*, *7*), thereby supporting a core finding in the study of American politics that the choice to be a Democrat or a Republican is largely shaped by parental socialization (*8*). However, other studies showed that the decision to affiliate with any political party (and the strength of this attachment) are significantly influenced by genes (*9*, *10*).

These initial twin studies suggested that political ideas are heritable, but they said little about political behavior. That changed this year, when a study (*11*) examined the heritability of voter participation by matching publicly available voter registration records to a twin registry in Los Angeles (*12*), analyzing self-reported turnout in the National Longitudinal Study of Adolescent Health (Add Health), and analyzing other forms of political participation. In all three cases, both genes and environment contributed significantly to variation in political participation (Fig. 1).

Other scholars wondered whether there might be similar variation in basic economic behavior. For example, they administered a "trust" game to twins in the United States and Sweden in which one (anonymous) subject decides how much to "invest" in another subject, the amount invested

is tripled by the researcher, and the recipient decides how much to return to the investor. They found that the behavior of both the investor and the recipient was significantly heritable, which suggests that genetic variation contributes to variation in cooperative behavior in the laboratory (*13*). Similar behavioral economics experiments have also indicated that altruism (*14*), bargaining (*15*), and attitudes toward risk (*14*) are heritable, and that variation in these traits plays an important role in political behavior (*16*, *17*).

These studies suggest that heritable factors are involved in political behavior and cooperation more generally, but they say nothing about which genes might play a role. Scholars therefore have turned their attention to specific genes that might be associated with political behaviors and attitudes, particularly those that affect the regulation of neurotransmitters. Dopamine and serotonin have been studied for several years and have been shown to influence social behavior in both animals and humans, so early work on politics has been directed at genes that affect their regulation (*18*, *19*). A direct association was recently established between voter turnout and the monoamine oxidase A (*MAOA*) gene, as well as a gene-environment interaction between turnout and the serotonin transporter (*5HTT*) gene, among those who frequently participated in religious activities (*18*). In other research, scholars have also found an association between voter turnout and a dopamine receptor (*DRD2*) gene that is mediated by a significant association between that gene and the tendency to affiliate with a political party (*19*). This work is preliminary and replication will be crucial, but it suggests that neurotransmitter function has an effect on political behavior. Future studies will also need to investigate whether genes influence political behavior predominantly through neurotransmitters and other cellular-level processes; through larger-scale differences in brain structure, function, or connectivity; through broader psychological constructs such as personality (*20*); or through a complex mix of all three (*21*).

Neurobiology and Politics

The genetic evidence so far has been about variation in political behavior, but there is also a stable core to this behavior that differentiates humans from other species. Synthesizing five decades of research, psychologists have recently identified a motivational basis for the stable, definitional core of conservative ideology, claiming that it is adopted in part to satisfy a variety of social, cognitive, and psychological needs (*22*).

Like our primate cousins, we are naturally adept at a variety of skills needed to navigate the everyday politics of our social species. However, although young humans appear to perform at a level similar to orangutans and chimpanzees in tasks involving technical problem solving, humans are far more sophisticated when it comes to social tasks (*23*).

It is easy to imagine that politics is just a cognitive exercise, like learning math or history in elementary school. However, neuroimaging evidence suggests that politics is not like subjects taught in the classroom. Instead, politics may be a form of social cognition that we have called "playground cognition" (*24*). On the playground, we are figuring out whom to cooperate with and whom to avoid; we are cognizant of social hierarchy; and we engage in coalitional cognition, knowing that an alliance with one group will entail exclusion from another. Even at rest on the playground, we are constantly monitoring our social environment and our place in it.

Neuroscientists have been studying a network of brain regions that diminishes in activity when subjects are engaged in a wide variety of technical cognitive tasks (*25*). One puzzle with this resting or "default state" network is that it consumes a large portion of the brain's metabolic budget and yet appears to deactivate under many conditions of active cognition. Meanwhile, when people make personal moral judgments (*26*) or observe social interactions (*27*), this network of brain regions increases in activity above the resting baseline. One component of this network, the medial prefrontal cortex, appears to be involved in thinking about the mental states of others. Another region, the medial parietal cortex and posterior cingulate, is active while we implicitly evaluate the social environment around us and is also implicated in emotional processing.

When people who are knowledgeable about national politics are asked for judgments of political issues or to attend to faces of national political figures, they increase the level of activity in the default state network above the resting baseline; such findings suggest that political thinking is akin to social cognition (Fig. 2) (*24*, *28*). Politically sophisticated subjects—both Republicans and Democrats—are using the same brain regions when they think about national politics. However, people who do not know much about national politics actually deactivate this set of brain regions, as if they had to treat these political questions as a form of technical cognition. We would not expect political novices to have some fundamental impairment on the playground. Instead, they appear to be merely unfamiliar with the specific domain of national politics. In contrast, people with autism spectrum disorders do appear to be generally unable to use their default state network properly (*29*, *30*), and although some of them are able to perform very well in the classroom, they struggle with the social cognition skills demanded on the playground.

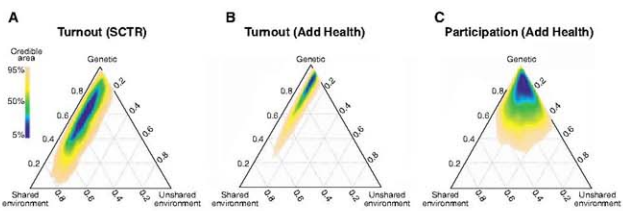


Fig. 1. Political participation is heritable. Ternary plots show estimates from a twin study model of **(A)** voter turnout among subjects in the Southern California Twin Registry (SCTR), **(B)** voter turnout among subjects in the National Longitudinal Study of Adolescent Health (Add Health), and **(C)** political participation among subjects in Add Health (an index that includes contributing money to a campaign, contacting a public official, running for office, or attending a rally or march). An additive genetic model uses identical and fraternal twin covariances to decompose the variance of a trait with respect to genetics, shared environment, and unshared environment factors. Colors indicate probabilities. The blue areas indicate the regions that are most likely to contain the true estimates; the beige areas indicate the region of 95% confidence (i.e., the probability that the true coefficients lie outside the colored regions is $P = 0.05$). Mean contribution of the genetic factor is estimated to be 53% for SCTR turnout, 72% for Add Health turnout, and 60% for Add Health participation (*11*).

The New Science of Human Nature

Large-scale political behavior is an extremely recent phenomenon in the span of human evolution, but the initial evidence suggests that it relies on genetic and neural mechanisms that evolved to solve basic social problems. These problems are inherently political because they involve decisions about the organization of humans to achieve group goals and the distribution

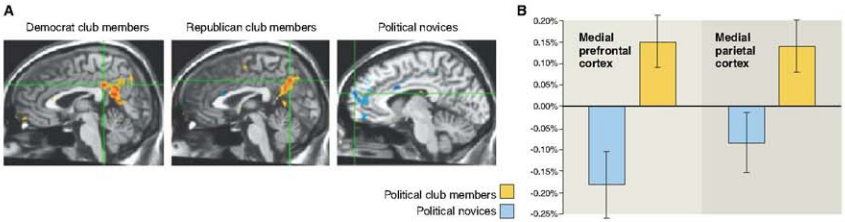


Fig. 2. Politics is a form of playground cognition. (**A** and **B**) When both college Democrat club members [(A), top left; cross-hairs at (2, −42, 33) with $z = 3.94$] and college Republican club members [(A), top right; cross-hairs at (1, −65, 37) with $z = 3.90$] answer questions about national politics, they demonstrate increases (colored orange and red) in the blood oxygen level–dependent functional magnetic resonance imaging (BOLD fMRI) signal above a resting baseline in the medial prefrontal cortex [(B), left; 4-mm spherical region of interest (ROI) centered at (−8, 59, 21)] and the medial parietal cortex [(B), right; 4-mm spherical ROI centered at (2, −64, 30)], brain regions that are part of a resting state network that has been shown to be active during social cognition. In contrast, college students with little knowledge of national politics [(A), bottom; cross-hairs at (−7, 61, 26) with $z = -3.55$] have a pattern of decreases (colored blue) in the same regions [(B), left and right]. Deactivations in this network typically occur during the kinds of technical cognition tasks performed in the classroom (*24*).

of resources within a group. But they are also inherently biological. For example, one of the most fundamental unanswered questions in evolutionary biology is how cooperative behavior evolved (*31*). If natural selection favors fit individuals, why do some individuals voluntarily reduce their fitness in order to enhance the fitness of others? Meanwhile, in political science we are focused on the nearly identical problem of collective action (*32*). In large-scale societies, why do people join political groups, participate in elections, and engage in other kinds of mass behavior when they know their efforts will not alter the political outcome?

Although simple forms of cooperation can be found far back in our evolutionary history, more sophisticated forms are quite recent. Evidence of political behavior in chimpanzees (*33*), capuchins (*34*), and early human societies (*35*) suggests that we may have, in part, evolved in ways that maximize our capabilities for small-scale interactions. But what are the biological mechanisms that enable us to cope with these small-scale interactions? And were new adaptations necessary for the development of large-scale political attitudes, behaviors, and institutions, or are we merely conserving other older tools for a new purpose?

Like Aristotle, we believe that the study of human nature should lie at the intersection of politics and biology. Whereas physiological studies on the correlation between political attitudes and biological factors can be rigorously performed on small numbers of subjects (*36*), political scientists are particularly adept at testing the implications of their theories in large populations (e.g., with surveys, natural experiments, and field experiments). Similarly, puzzles that arise at the population level can drive the kind of laboratory work that neuroscientists are more familiar with (*37*).

When Aristotle wanted to understand how humans govern themselves, he started by cataloging political institutions. Today, the study of institutions has improved our understanding of political outcomes because they help us understand how legislatures, courts, and other bodies are constrained in their behavior. Similarly, the study of genes potentially promises a better understanding of the constraints imposed on basic political psychology. The new science of human nature demands that we recognize that genes are the institutions of the human body. They regulate the neurological processes that drive social and political behavior. And we cannot fully appreciate their function in humans without understanding their role in the very complex social and political interactions that characterize our species.

References

1. Aristotle, in *Cambridge Texts in the History of Political Thought*, S. Everson, Ed. (Cambridge Univ. Press, Cambridge, 1996).
2. R. W. Byrne, A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans* (Oxford Univ. Press, Oxford, 1988).
3. J. G. Matsusaka, F. Palda, *Public Choice* **98**, 431 (1999).
4. E. Plutzer, *Am. Polit. Sci. Rev.* **96**, 41 (2002).
5. N. G. Martin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4364 (1986).
6. J. R. Alford, C. L. Funk, J. R. Hibbing, *Am. Polit. Sci. Rev.* **99**, 153 (2008).
7. P. K. Hatemi, S. E. Medland, K. I. Morley, A. C. Heath, N. G. Martin, *Behav. Genet.* **37**, 435 (2007).
8. A. Campbell, P. E. Converse, W. E. Miller, D. E. Stokes, *The American Voter* (Univ. of Chicago Press, Chicago, 1960).
9. J. E. Settle, C. T. Dawes, J. H. Fowler, *Polit. Res. Q.*, in press.
10. P. K. Hatemi, J. Hibbing, J. Alford, N. Martin, L. Eaves, *Polit. Res. Q.*, in press.
11. J. H. Fowler, L. A. Baker, C. T. Dawes, *Am. Polit. Sci. Rev.* **102**, 233 (2008).
12. L. A. Baker, M. Barton, A. Raine, J. H. Fowler, *Twin Res. Hum. Genet.* **9**, 933 (2006).
13. D. Cesarini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3721 (2008).
14. D. Cesarini, C. T. Dawes, M. Johannesson, P. Lichtenstein, B. Wallace, *Q. J. Econ.*, in press.
15. B. Wallace, D. Cesarini, P. Lichtenstein, M. Johannesson, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15631 (2007).
16. J. H. Fowler, *J. Polit.* **68**, 674 (2006).
17. J. H. Fowler, C. D. Kam, *J. Polit.* **69**, 813 (2007).
18. J. H. Fowler, C. T. Dawes, *J. Polit.* **70**, 579 (2008).
19. C. T. Dawes, J. H. Fowler, *J. Polit.*, in press.
20. J. M. Olson, P. A. Vernon, J. A. Harris, K. L. Jang, *J. Pers. Soc. Psychol.* **80**, 845 (2001).
21. N. I. Eisenberger, B. M. Way, S. E. Taylor, W. T. Welch, M. D. Lieberman, *Biol. Psychiatry* **61**, 1100 (2007).
22. J. Jost, *Am. Psychol.* **61**, 651 (2006).
23. E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, M. Tomasello, *Science* **317**, 1360 (2007).
24. D. Schreiber, in *The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior*, A. Crigler, M. MacKuen, G. E. Marcus, W. R. Neuman, Eds. (Univ. of Chicago Press, Chicago, 2007), pp. 48–70.
25. M. E. Raichle *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 676 (2001).
26. J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, J. D. Cohen, *Science* **293**, 2105 (2001).
27. M. Iacoboni *et al.*, *Neuroimage* **21**, 1167 (2004).
28. D. Westen, P. Blagov, K. Harenski, C. Kilts, S. Hamann, *J. Cogn. Neurosci.* **18**, 1947 (2006).
29. M. Iacoboni, *Trends Cognit. Sci.* **10**, 431 (2006).
30. D. P. Kennedy, E. Redcay, E. Courchesne, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8275 (2006).
31. R. M. May, in A. M. Colman, *Nature* **440**, 744 (2006).
32. E. Ostrom, *Am. Polit. Sci. Rev.* **92**, 1 (1998).
33. F. B. M. de Waal, *Chimpanzee Politics: Power and Sex Among Apes* (Johns Hopkins Univ. Press, Baltimore, 1982).
34. S. F. Brosnan, C. Freeman, F. B. M. de Waal, *Am. J. Primatol.* **68**, 713 (2006).
35. C. Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard Univ. Press, Cambridge, MA, 1999).
36. D. M. Amodio, J. T. Jost, S. L. Master, C. M. Yee, *Nat. Neurosci.* **10**, 1246 (2007).
37. M. Lieberman, D. Schreiber, K. Ochsner, *Polit. Psychol.* **24**, 681 (2003).

10.1126/science.1158188

Bioactive Contaminants Leach from Disposable Laboratory Plasticware

G. Reid McDonald,[1] Alan L. Hudson,[1] Susan M. J. Dunn,[1] Haitao You,[1] Glen B. Baker,[2] Randy M. Whittal,[3] Jonathan W. Martin,[4] Amitabh Jha,[5] Dale E. Edmondson,[6] Andrew Holt[1]*

Recent reports of leaching of bisphenol A and antimony into foods and beverages from polycarbonate and polyethylene terephthalate containers, respectively, have drawn attention to plastics as potential sources of bioactive environmental contaminants (*1*, *2*). However, numerous other processing additives coat, or intercalate within, polymeric structures, and these also migrate into foods stored in plastic containers (*3*).

Disposable plasticware is used in life science laboratories worldwide. Although labeling of plastics as "sterile" appears to offer researchers some assurance that products are free of bioactive contaminants, the presence of processing additives is unavoidable. Herein, we report identification of two additives leaching from disposable plasticware and demonstrate potent effects on enzyme and receptor proteins.

Observations of anomalous kinetics with human monoamine oxidase-B (hMAO-B), which recognizes numerous xenobiotic substrates, led us to examine disposable tubes used in our assays as a potential source of interferences. Water rinsed through several brands and sizes of plastic tubes adopted inhibitory potency versus hMAO-B; when dimethyl sulfoxide (DMSO) (10%, v/v) was used instead, inhibition was more pronounced, and marked enzyme activation was observed in one case (Fig. 1A). Samples of two tubes (indicated) were rinsed with water (W) or methanol (M), and leachates were dried and subjected to mass spectrometry. Fragmentation spectra for major species present (Fig. 1B and fig. S1A) identified a biocide, di(2-hydroxyethyl)methyldodecylammonium (DiHEMDA), and a slip agent, 9-octadecenamide (oleamide), in W and M leachates, respectively (Fig. 1C, inset). Pure samples of DiHEMDA (*4*) and oleamide were confirmed as hMAO inhibitors, showing selectivity for hMAO-B (Fig. 1C). Inhibition by leachates from oleamide-positive tubes increased markedly over a 10-day plastic exposure period at 20°C (fig. S1B).

Repeated (10 times) pipetting of 10% DMSO with pipette tips from several suppliers resulted in extraction of species that had significant effects on hMAO-B (fig. S1C). One or more unidentified compounds causing hMAO-B activation also leached into 10% DMSO from wells of polystyrene or acrylic 96-well microplates (fig. S1D).

Oleamide is an endogenous signaling molecule that binds to numerous receptor and channel proteins (*5*), including the γ-aminobutyric acid type A ($GABA_A$) receptor. Specific binding of the $GABA_A$ radioligand [^{3}H]Ro15-4513 to rat brain membranes was inhibited significantly by DMSO leachate from Eppendorf (Eppendorf AG, Hamburg, Germany) microfuge tubes (Fig. 1D).

Related slip agents such as erucamide and stearamide are endogenous molecules used routinely in plastic manufacturing (*3*), whereas quaternary ammonium compounds are included as biocides or antistatic agents. Many such biocides bind substantially to proteins and DNA and have recently been linked with fertility problems in mice (*6*). Our findings that processing agents leach from laboratory plasticware into biological media and solvents, particularly when liquids are stored in plastic vessels, identify a likely source of error in many assay systems.

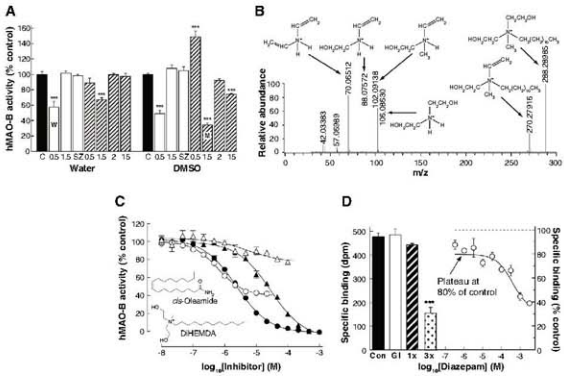


Fig. 1. **(A)** Effects on hMAO-B of water ($N = 3$) or 10% DMSO ($N = 7$) (40% of tube volume, 1 hour, 20°C) from Fisherbrand (Fisher Scientific, Ottawa, Canada) (clear bars) or Sarstedt (Sarstedt AG, Nümbrecht, Germany) (hatched bars) plastic tubes or glass vials (C). *x* axis labels indicate tube volumes (ml). C indicates control; SZ, siliconized 0.5-ml tube; W and M, water and methanol rinses analyzed by tandem mass spectrometry (MS/MS). Data are mean ± SEM. ***$P < 0.001$ compared with control (one-way analysis of variance with Dunnett's post hoc test). **(B)** MS/MS data for the major peak obtained from a water rinse of a Fisherbrand 0.5-ml tube reveal the presence of DiHEMDA. **(C)** Inhibition of human MAO-A (triangles) or MAO-B (circles) by oleamide (open symbols) or DiHEMDA (solid symbols) (structures inset). **(D)** (Left) Effects on binding of [^{3}H]Ro15-4513 to rat brain $GABA_A$ channels of DMSO (0.2% in assay) rinsed (at 100%) through a glass (GI) tube or through one (1×) or three consecutive (3×) Eppendorf brand 1.5-ml tubes, compared with a DMSO-free control (Con). (Right) Competition binding curve to diazepam, prepared in DMSO in microfuge tubes, reveals a lowered plateau relative to a DMSO (in glass) control.

References and Notes

1. L. N. Vandenberg, R. Hauser, M. Marcus, N. Olea, W. V. Welshons, *Reprod. Toxicol.* **24**, 139 (2007).
2. W. Shotyk, M. Krachler, *Environ. Sci. Technol.* **41**, 1560 (2007).
3. I. Cooper, P. A. Tice, *Food Addit. Contam.* **12**, 235 (1995).
4. Details, including methods, are available as supporting material on *Science* Online.
5. C. R. Hiley, P. M. Hoi, *Cardiovasc. Drug Rev.* **25**, 46 (2007).
6. Q&A News Article, *Nature* **453**, 964 (2008).
7. Supported by the Canadian Institutes of Health Research (A.H., S.M.D., and G.B.B.), the Canada Foundation for Innovation (G.B.B.), the Natural Sciences and Engineering Research Council of Canada (A.J.), and the National Institute of Health (D.E.E.). We thank O. Degenhardt, I. Paulsen, and G. Rauw for technical assistance.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/917/DC1
Materials and Methods
Fig. S1
References

26 June 2008; accepted 2 September 2008
10.1126/science.1162395

[1]Department of Pharmacology, University of Alberta, Edmonton, AB T6G 2B7, Canada. [2]Department of Psychiatry, University of Alberta, Edmonton, AB T6G 2R7, Canada. [3]Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada. [4]Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB T6G 2G3, Canada. [5]Department of Chemistry, Acadia University, Wolfville, NS B4P 2R6, Canada. [6]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA.

*To whom correspondence should be addressed. E-mail: aholt@pmcol.ualberta.ca

Global Protein Stability Profiling in Mammalian Cells

Hsueh-Chi Sherry Yen, Qikai Xu,* Danny M. Chou,* Zhenming Zhao, Stephen J. Elledge†

The abundance of cellular proteins is determined largely by the rate of transcription and translation coupled with the stability of individual proteins. Although we know a great deal about global transcript abundance, little is known about global protein stability. We present a highly parallel multiplexing strategy to monitor protein turnover on a global scale by coupling flow cytometry with microarray technology to track the stability of individual proteins within a complex mixture. We demonstrated the feasibility of this approach by measuring the stability of ~8000 human proteins and identifying proteasome substrates. The technology provides a general platform for proteome-scale analysis of protein turnover under various physiological and disease conditions.

A complete understanding of biological networks requires knowledge of all aspects of regulation, from gene transcription, RNA processing, translation, and localization to protein modification and turnover. However, the chemical heterogeneity of proteins, the large dynamic range of their abundance, and the absence of specific recognition reagents have hindered high-throughput systematic approaches for the analysis of protein regulation (*1*, *2*). Thus, protein stability remains an unexploited feature for most proteins (*3*).

The selective degradation of proteins is critical for most cellular processes, including cell cycle progression, signal transduction, and differentiation (*4*–*6*). Control of protein turnover serves as a rapid mechanism for activation or inhibition of signaling pathways when cells respond to environmental changes. Alterations in the degradation of cancer-related proteins have important roles in cellular transformation, and multiple components of the proteolysis system are directly involved in human diseases (*7*–*9*). Furthermore, many viruses have evolved strategies to hijack the proteolytic pathway of host cells for their own benefit (*10*). Therefore, understanding protein turnover should have a major impact, not only for the basic science of biological regulation, but also for the development of more effective strategies to cure diseases (*11*, *12*).

Traditional methods of measuring protein stability rely on either pulse-chase metabolic labeling or administration of protein synthesis inhibitors, followed by biochemical analysis of the abundance of the protein of interest at multiple time points during the chase period. When applying half-life analysis to a global population of proteins under a broad range of physiological or disease states, these assays are impractical. One method to increase throughput is to combine pulse-chase analysis with mass spectrometry (*3*). However, mass spectrometry is limited by the inability to detect low-abundance proteins, and highly regulated proteins tend to be present in low amounts. Nor can these methods be used for real-time monitoring of protein turnover in living cells with single-cell resolution, a feature important for systems level understanding of protein function (*13*–*15*). Therefore, we developed a high-throughput approach for proteome-scale protein-turnover analysis in mammalian cells, called global protein stability (GPS) analysis and described here, that overcomes many of these deficiencies (*16*).

Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: selledge@genetics.med.harvard.edu

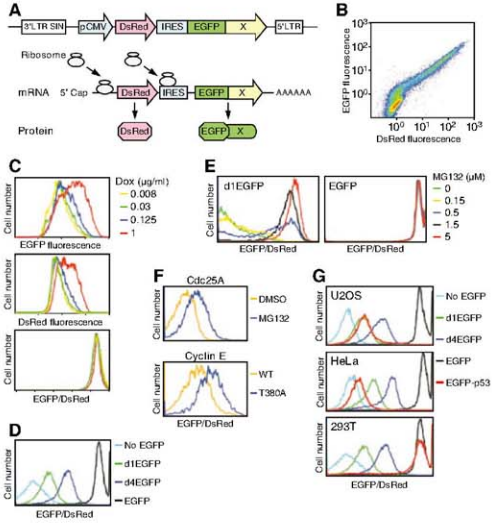


Fig. 1. Determination of protein stability by the GPS system. **(A)** A schematic representation of the reporter construct. Ribosomes can dock at the 5′ end of the bicistronic mRNA to translate DsRed or at the IRES to translate EGFP-X. **(B)** HEK 293T cells stably expressing the pCMV-DsRed-IRES-EGFP reporter were analyzed by FACS. The cluster in the lower left corner of the plot represents uninfected cells. **(C)** HeLa tetracycline-inducible (tet-on) cells with the pTRE-DsRed-IRES-EGFP expression cassette were treated with various doses of Dox and analyzed by FACS. **(D)** HEK 293T cells carrying the pCMV-DsRed-IRES-EGFP reporter cassette with EGFP, d4EGFP, or d1EGFP or without EGFP were analyzed by FACS. **(E)** HEK 293T cells expressing d1EGFP or EGFP from the DsRed-IRES-EGFP reporter cassette were treated with various concentrations of MG132 and analyzed. **(F)** (Top) HEK 293T cells expressing the DsRed-IRES-EGFP-Cdc25A reporter with or without MG132 treatment (2 μM) were compared. (Bottom) HEK 293T cells with the reporter carrying EGFP-fused wild-type or T380A mutant cyclin E were analyzed. **(G)** Overlay of the EGFP/DsRed ratios of EGFP-p53 and various EGFP-degron fusions in the three different cell lines, as indicated in the graphs.

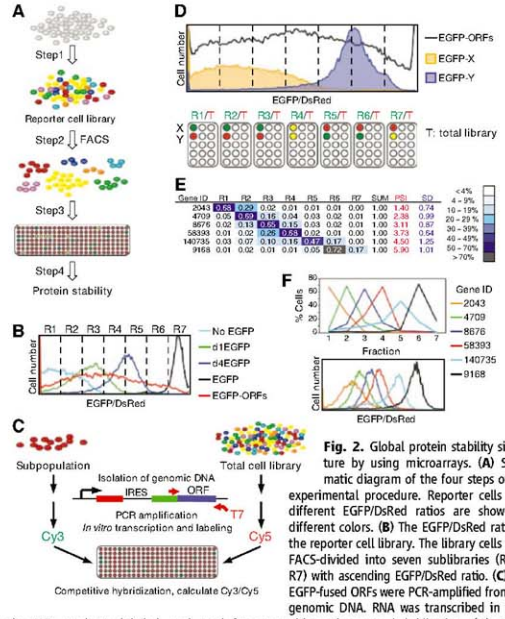


Fig. 2. Global protein stability signature by using microarrays. **(A)** Schematic diagram of the four steps of the experimental procedure. Reporter cells with different EGFP/DsRed ratios are shown in different colors. **(B)** The EGFP/DsRed ratio of the reporter cell library. The library cells were FACS-divided into seven sublibraries (R1 to R7) with ascending EGFP/DsRed ratio. **(C)** The EGFP-fused ORFs were PCR-amplified from the genomic DNA. RNA was transcribed in vitro from the PCR products, labeled, and used for competitive microarray hybridization. **(D)** Seven hybridizations were performed with ORFs from the total library labeled with Cy5 and ORFs from each FACS-isolated subpopulation labeled with Cy3. The EGFP/DsRed ratio of cells expressing EGFP fused to a protein of interest can be tracked by combining the hybridization results (Cy3/Cy5) from the seven microarrays. Two EGFP fusion proteins, EGFP-X (short half-life) and EGFP-Y (long half-life) are shown as examples. **(E)** Representative data. **(F)** Results for individual ORFs from microarrays [(top) percent of cells in each fraction] and FACS analysis [(bottom) EGFP/DsRed ratio] were compared.

A fluorescence-based system to monitor protein stability at the single-cell level. We established a cell-based system for measuring GPS. We built a retroviral reporter construct in which the expression cassette contains a single promoter that, with an internal ribosome entry site (IRES) (*17*, *18*), permits the translation of two fluorescent proteins from one mRNA transcript (Fig. 1A). The first fluorescent protein, *Discosoma* sp. red fluorescent protein (DsRed), served as an internal control, whereas the second, enhanced green fluorescence protein (EGFP), was expressed as a fusion with the protein of interest (X). When integrated into the genome of cells, DsRed and EGFP-X proteins should be produced at a constant ratio because they are derived from the same mRNA (Fig. 1A). The EGFP/DsRed ratio of cells represents the stability of protein X in this system and can be quantified by fluorescence-activated cell sorting (FACS). Events that selectively affect the protein stability of EGFP-X (for example, depletion or overexpression of proteins that regulate X turnover) are expected to change the abundance of EGFP-X, but not DsRed, and thus lead to an alteration of the EGFP/DsRed ratio.

The GPS system has several important features. Because the EGFP and DsRed proteins are derived from the same mRNA, the EGFP/DsRed ratio is not affected by transcriptional regulation. In addition, the reporter construct is integrated into the genome as a single copy, and EGFP-X is expressed downstream of an IRES under the control of a ubiquitously active promoter. Because translation of the gene downstream of an IRES is less efficient than that of the upstream gene (*17*), this design limits overproduction of the EGFP-X protein and potential phenotypic perturbation in cells. Using a common promoter to drive reporter protein expression also reduces the range of protein abundance, a common problem for global proteomic assays. In contrast to traditional protein half-life assays that measure only the mean value of protein half-lives from a population of cells,

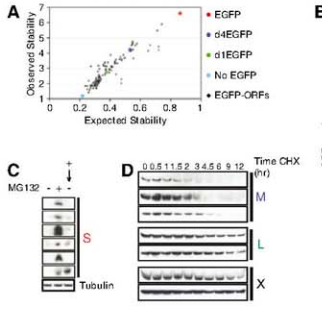


Fig. 3. Screen validation. **(A)** Unbiased validation of microarray performance. The expected stability is derived from the EGFP/DsRed ratio from individual FACS analysis; the observed stability is the PSI derived from microarray results. The expected and observed stabilities were compared for 96 ORFs from ORFeome plate 11001. **(B)** The steady-state protein levels of 75 randomly chosen ORF-HA fusions were analyzed by Western blotting and quantified. Additional information and data related to the tested ORFs can be found in table S5. The number of ORFs with a particular range of expression was counted and plotted on a graph. S, M, L, and X represent proteins with short, medium, long, and extra-long half-lives, according to the PSI. **(C)** HEK 293T cells expressing short–half-life ORF-HA fusions were either left untreated (lane 1), treated with 2 μM MG132 for 2 hours (lane 2), or treated with MG132 for 2 hours and released for 6 hours (lane 3), and analyzed by Western blot. **(D)** Cycloheximide (CHX)–chase analysis of cells expressing medium, long, and extra-long–half-life ORF-HA proteins. Protein samples were normalized based on cell number.

the GPS system allows real-time protein stability detection at the level of individual living cells. Last, this fluorescence-based system has the potential for automation and is amenable to high-throughput proteome-scale analysis.

We integrated the GPS reporter that expresses DsRed and EGFP under the control of the cytomegalovirus (CMV) promoter and the IRES from encephalomyocarditis virus into the genome of human embryonic kidney (HEK) 293T cells and analyzed cells by FACS. Depending on the integration site, amounts of the fluorescent proteins might differ in particular cells, but the EGFP/DsRed ratio should be constant. As expected, the EGFP and DsRed signals displayed a constant ratio despite the wide range of fluorescence intensities for each protein in different cells (Fig. 1B), which suggests that this ratio is independent of transcription. To confirm this, we replaced the CMV promoter with the tetracycline response element (TRE) promoter that is controllable by doxycycline (Dox). The EGFP/DsRed ratio did not vary, despite increases in EGFP and DsRed abundance after the addition of Dox (Fig. 1C). To test whether the EGFP/DsRed ratio can be used to distinguish proteins with different half-lives, we created cell lines that express DsRed-IRES-EGFP fusions with mutants of the ornithine decarboxylase (ODC) degron that confer different half-lives (d1EGFP, $t_{1/2}$ = 1 hour; d4EGFP, $t_{1/2}$ = 4 hours; EGFP, $t_{1/2}$ = 24 hours). Cells stably expressing these fusions displayed distinct EGFP/DsRed ratios that reflect the stability of the EGFP fusions (Fig. 1D). We obtained similar results in all cell types tested (Fig. 1G). These data indicate that the EGFP/DsRed ratio is a reliable readout for protein stability in mammalian cells.

Comparison of the EGFP/DsRed ratio to other measurements of protein stability. We examined the correlation between our protein stability measurements and those reported in the literature. The ODC degron destabilizes EGFP by targeting it for hydrolysis by the proteasome (*19*). Consistent with that, the EGFP/DsRed ratio of cells expressing EGFP to which the degron was fused increased in response to treatment with the proteasome inhibitor MG132 in a dose-dependent manner (Fig. 1E). We also made reporter cells expressing EGFP-Cdc25A, EGFP–cyclin E and EGFP-p53 fusions. The stability of Cdc25A increased in cells treated with MG132, and cyclin E harboring the T380A mutation (substitution of Ala for Thr^{380}) displayed greater stability than the wild-type protein (*20*) (Fig. 1F). These results are consistent with the previous discoveries that Cdc25A is a proteasome substrate and that mutation of Thr^{380} in cyclin E results in defective turnover (*21–24*).

We measured the half-life of p53 in three different cell lines with distinct p53 turnover pathways. Comparison of the EGFP/DsRed ratio of EGFP-p53–expressing cells with that of the EGFP-degron series indicated that p53 displays distinct stabilities in these three cell lines (Fig. 1G). In human osteosarcoma U2OS cells, the half-life of EGFP-p53 approaches that of d1EGFP. HeLa cells, which express papillomavirus E6, promote more rapid p53 degradation through E6-associated protein (E6-AP) (*25*) and had a lower EGFP/DsRed ratio. In contrast, EGFP-p53 is highly stable in HEK 293T cells, which express the T antigen that binds and sequesters p53 (*26*). In conclusion, the EGFP/DsRed ratio serves as an accurate measurement of the relative half-lives of proteins and is responsive to genetic changes.

Multiplex GPS profiling by using DNA microarray deconvolution. To apply GPS on a global scale, we designed a multiplexing strategy that integrates the power of fluorescence-based protein stability analysis with that of DNA microarray technology to rapidly obtain protein stability profiles in human cells. The four elements of this approach are shown in Fig. 2A. To create a reporter cell library, we generated a pCMV-DsRed-IRES-EGFP-fusion cDNA library in a retroviral vector using the hORFeome v1.1 library, which consists of ~8000 unique, full-length human protein–encoding open reading frames (ORFs) in a Gateway entry vector that allows in-frame transfer of the ORFs to any expression plasmid by recombinational cloning (*27*). The hORFeome library gene set is an arrayed set of individual ORFs in 96-well plates, which facilitates future screen validation processes. We used the retroviral DsRed-IRES-EGFP-ORF library to produce viruses and infected HEK 293T cells to make DsRed-IRES-EGFP-ORF v1.1 reporter cell collections. To ensure that each cell carried only one reporter cassette, cells were infected at low multiplicity of infection (MOI ≈ 0.05).

We then fractionated the cell library into seven subpopulations of increasing EGFP/DsRed ratios (R1 to R7) by FACS (Fig. 2B). Because sorting of cells into distinct pools was dependent on the half-lives of the EGFP-ORF proteins they express, the stability of a particular EGFP-ORF fusion could be inferred from the distribution of cells expressing that specific fusion within the seven pools. To determine the protein stability of all 8000 ORFs, the EGFP-fused ORF sequences that serve as cell identifiers were amplified by the polymerase chain

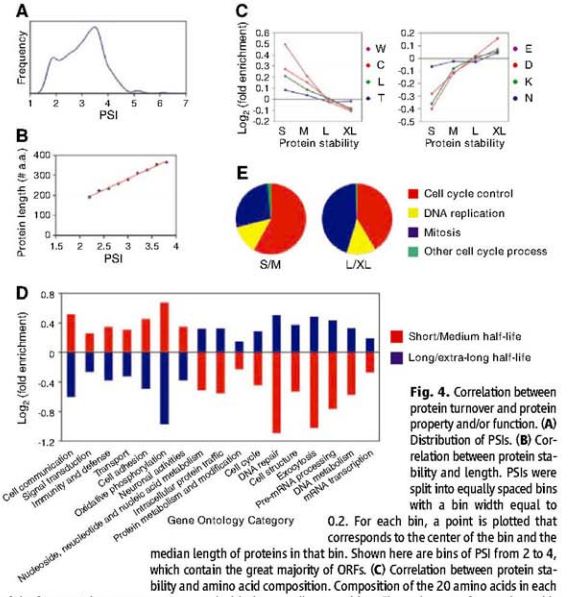


Fig. 4. Correlation between protein turnover and protein property and/or function. **(A)** Distribution of PSIs. **(B)** Correlation between protein stability and length. PSIs were split into equally spaced bins with a bin width equal to 0.2. For each bin, a point is plotted that corresponds to the center of the bin and the median length of proteins in that bin. Shown here are bins of PSI from 2 to 4, which contain the great majority of ORFs. **(C)** Correlation between protein stability and amino acid composition. Composition of the 20 amino acids in each of the four protein groups was compared with the overall composition. Shown here are four amino acids that are enriched in short–half-life proteins (left) and four that are enriched in long–half-life proteins (right). **(D)** Correlation between protein stability and function. Presented are Panther Biological Process GO categories that are inversely enriched between the short–/medium–half-life proteins and long–/extra-long–half-life proteins. **(E)** Distribution of short–/medium– (left) and long–/extra-long–half-life (right) cell cycle proteins in "cell cycle" subcategories.

reaction (PCR) from the genomic DNA of both the sorted subpopulations and the total library, and their relative abundance was quantified in microarrays (Fig. 2C). Unlike transcription profiling, which uses oligonucleotides from "mRNA" for hybridization, the ORF sequences are recovered from the "genomic DNA" of reporter cells; therefore, the microarray result reflects the composition of the library and is not affected by transcription. Because normal transcriptional microarrays are biased for probes in the 3′ untranslated region, we designed our own custom microarray to specifically detect ORF sequences (table S1).

We performed seven hybridizations of the sorted subpopulation of cells (Cy3-labeled) versus the total library (Cy5-labeled) and combined seven sets of array data to determine the abundance of different ORFs in the various pools (Fig. 2D and fig. S1A). All cells expressing a specific EGFP-ORF fusion should be distributed within the seven subpopulations after FACS, and thus, the sum of the Cy3/Cy5 ratios from seven hybridizations for a probe should be equal to 1 (100%). To represent the protein stability information from arrays in a quantitative manner, we calculated the "protein stability index" (PSI) using the following formula: $\mathrm{PSI} = \sum_{i=1}^{7} R_i * i$, where i is the subpopulation number (that is 1 to 7) and R_i is the fraction of the signal present for a gene in that given subpopulation i. The value of PSI ranges from 1 to 7, with a higher PSI value meaning higher relative protein stability of EGFP-fused ORFs. We also calculated the standard deviation $[\mathrm{SD} = \sqrt{\sum_{i=1}^{7} R_i * (i - \mathrm{PSI})^2}]$ to quantify the spread of protein stabilities in a population of cells expressing a specific EGFP-ORF fusion. The complete list of screen results is in table S2, and representative data are shown in Fig. 2E.

Screen results and validation. We carried out a series of analyses and confirmed that the overall quality of the array hybridization is high (*16*) (fig. S2 and tables S3 and S4). Our data indicate that >98% of ORFs from the hORFeome v1.1 library are preserved in the reporter cell library. To examine whether the protein stability information derived from the microarray data resembled that measured by FACS analysis of individual samples, we randomly picked one plate (96 ORF clones) of the arrayed ORFeome collection for analysis. Each ORF was individually recombined into the DsRed-IRES-EGFP vector, packaged into viruses, and transduced independently into cells. The EGFP/DsRed ratio for each clone was measured by FACS and compared with the PSI calculated from the array. There was a nearly perfect correlation between the "observed stability," which reflects PSI, and the "expected stability" derived from the EGFP/DsRed ratio of individual clones, with a correlation coefficient of 0.907 (Fig. 3A). Moreover, the distribution of EGFP/DsRed ratios for a given clone derived from the array was similar to that from individual FACS (Fig. 2F). This is striking because the PSI was derived from library cells divided into only seven pools. In addition, the PSI can be affected by double integrations, cell purity after FACS fractionation, ORF preparation, or hybridization differences. In summary, these data indicate that microarray deconvolution provides a robust readout for protein stability.

One potential caveat of this technology is that the N-terminal EGFP fusion might affect protein turnover in individual cases. To address this concern, we randomly picked 75 ORFs, tagged them with a single hemagglutinin (HA) epitope at the C terminus and expressed them under the control of the elongation factor 1α (EF1α) promoter at single copy. These ORFs belong to four stability categories on the basis of our array information: short (PSI ~ 1.3), medium (~3.5), long (~5.5), and extra-long (~6.5) (table S5). We found that the abundance of ORF-HA proteins determined by Western blot analysis correlated well with the PSI of EGFP-ORF fusions measured by microarrays (Fig. 3B and fig. S3). For example, only 1 out of 19 HA-fused proteins in the short–half-life category was detected, and that amount was low. In contrast, all 18 proteins from the extra-long–half-life category were detected, and most displayed strong Western signals. The abundance of a given cellular protein is determined by the balance between its rate of synthesis and degradation, so differences in the steady-state abundance of ORF-HA fusions are likely due to differences in their turnover rates. To ensure the inability to detect ORF-HA proteins with low PSIs was indeed due to degradation, we treated cells with MG132 and observed rapid accumulation of those proteins (Fig. 3C). We also treated cells for a limited time with the protein synthesis inhibitor cycloheximide to measure the stability of medium–, long–, and extra-long–half-life proteins and found that the PSI serves as an accurate indicator of protein stability (Fig. 3D). Thus, in general, the stabilities of most proteins are not compromised by the N-terminal EGFP tag and the gateway *att* sites.

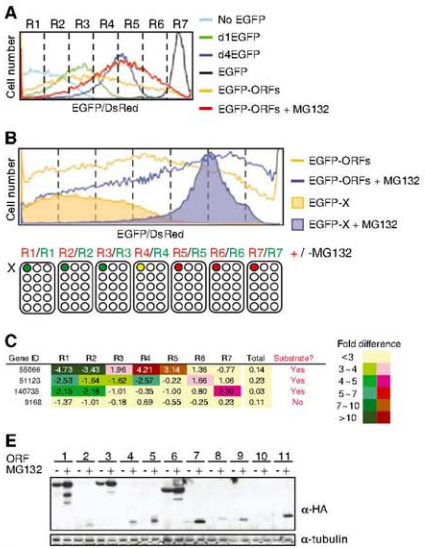


Fig. 5. Identification of proteasome substrates by comparative protein-stability profiling. **(A)** The EGFP/DsRed ratio of library cells with or without MG132 treatment. **(B)** Library cells with or without MG132 were fractionated into seven pools (R1 to R7). Seven hybridizations with untreated/Cy3 versus treated/Cy5 were performed, and data were combined. Data from cells expressing a proteasome substrate, X, are shown as an example. **(C)** Representative results. Numbers shown are $\log_2$(Cy5/Cy3). **(D)** Representative comparison between protein stability inferred from individual EGFP/DsRed FACS measurements (top) and that from microarray data (bottom). The ORFs shown are the same as those in (C). **(E)** The steady-state amount of ORF-HA proteins from cells treated with or without MG132 was analyzed by Western blot. ORF-HA fusion proteins were expressed from the EF1α promoter.

Protein stability and protein properties/functions. We used our protein stability information to investigate whether correlations exist between particular properties of proteins and protein half-lives. To clearly separate proteins with different half-lives, we filtered out genes with a flat or bimodal stability profile and kept only 5341 ORFs that had a single sharp peak for further bioinformatics analysis (*16*). The global distribution of PSIs showed a bimodal pattern, with a major peak centered around 3.5 and a minor peak centered at 2 (Fig. 4A). The PSIs for d1EGFP, d4EGFP, and EGFP were 2.75, 4.19, and 6.61, respectively. With a simple linear-regression analysis based on the PSI of these three known half-life proteins, we estimated that PSIs of 2 and 3.5 approximate protein half-lives of 30 min and 2 hours, respectively. There is a significant positive correlation between protein half-life and protein length [Spearman correlation coefficient, $r = 0.28$, $P < 10^{-16}$ (Fig. 4B)], which indicated that longer proteins tend to be more stable.

To survey whether amino acid composition of proteins influences turnover, we divided ORFs into four groups according to PSI: short half-life (S: PSI < 2; 653 ORFs), medium half-life (M: 2 ≤ PSI < 3; 1734 ORFs), long half-life (L: 3 ≤ PSI < 4; 2442 ORFs), and extra-long half-life (XL: PSI ≥ 4; 512 ORFs). We observed that the amino acids W, C, L, T, F, W, Y, and V were enriched in labile proteins, whereas charged amino acids including E, D, K, N, R, and Q were enriched in stable proteins (*20*) (Fig. 4C). We searched for enrichment of proteins with shared Gene Ontology (GO) terms in the four half-life groups to look for correlations between protein stability and function. Many shared categories were detected between short– and medium–half-life proteins, and many categories were shared between long and extra-long–half-life proteins. By contrast, there was no overlap of enriched GO categories between short–/medium– and long–/extra-long–half-life proteins. Short–half-life proteins are enriched for membrane proteins and signal transduction proteins, whereas long–half-life proteins are enriched for cytoskeleton proteins and nuclear proteins with housekeeping functions (Fig. 4D). We also examined the relation between protein stability and cell cycle subfunctional categories, and found that the short–/medium–half-life group has a larger proportion of "cell cycle control" proteins that are generally known to be unstable; the long–half-life group has a larger fraction of "mitosis" proteins that consist of actins, tubulins, septins, and so on (Fig. 4E). Because the N-terminal EGFP tag interferes with the signal peptide for membrane anchoring, it is possible that the rapid degradation of EGFP-fused membrane proteins was caused by their mislocalization. We think that this is not the case because 14 C-terminally HA-tagged proteins with short half-lives that were tested in Fig. 3 are membrane proteins, and all were unstable regardless of the position of the tag. We also wondered whether the decreased proportion of charged amino acids in proteins with short half-life reflected the hydrophobic nature of membrane proteins, but we obtained the same conclusion after removing all annotated membrane proteins from the data set used for the analyses.

To tightly and efficiently control the amount of proteins in cells, it is likely that transcription, translation, and protein turnover must be coordinated. Our data indicate a statistically significant positive correlation between the steady-state mRNA level and protein stability [Pearson correlation coefficient $r = 0.124$, $P < 2 \times 10^{-13}$ (fig. S4)]. A previous study in yeast also showed that low mRNA abundance correlates with instability (*28*).

Global identification of proteasome substrates by comparative protein stability profiling. A key goal of modern biological analyses is to determine how proteomes are remodeled in response to environmental cues. We developed a "comparative GPS profiling" strategy to identify proteins whose stabilities increased or decreased in response to stimuli. We used two pools of library cells, treated one with the stimulus of choice and sorted cells from each pool by FACS on the basis of EGFP/DsRed ratios. The relative clone distribution was deconvoluted by quantitative comparative hybridization. By comparing stability profiles from treated and untreated cells, we determined which proteins' stabilities were altered in response to the stimulus (fig. S1B).

To screen for proteasome targets, we treated library cells with MG132. EGFP itself is not degraded by the proteasome (Fig. 1E). When looking at the EGFP/DsRed profile of the entire cell library, we detected a global increase in the EGFP/DsRed ratio in response to MG132, which suggested that large numbers of ORFs encode proteasome substrates (Fig. 5A). We fractionated MG132-treated and untreated library cells into seven pools and isolated EGFP-fused ORFs for microarray hybridization. We hybridized probes derived from untreated (Cy3-labeled) versus treated (Cy5-labeled) fractions, as well as untreated versus treated total cell library, and determined the Cy5/Cy3 ratio distribution of probes for each chip (Fig. 5B). In agreement with the hypothesis that many short–half-life EGFP-ORFs might encode proteosome substrates, we observed the reduction of ORFs (Cy5 < Cy3) from cell fractions that contained proteins with short half-lives (R1 to R3) and an increase of ORFs (Cy5 > Cy3) in fractions that express proteins with higher stability (R4 to R7) after MG132 treatment. The Cy5 and Cy3 signals of most probes (95.28%) from the total library differed by less than 50%, which indicated that the overall EGFP-ORF compositions of untreated and MG132-treated cells were very similar (fig. S5). Therefore, any changes in the Cy5/Cy3 ratios represent a redistribution as opposed to overall gain or loss of signals.

We integrated and arrayed the $\log_2$(Cy5/Cy3) ratio for each probe from seven hybridizations (R1 to R7) to search for proteasome substrates. Cells expressing an EGFP-ORF that is not a proteasome target should be sorted into the same EGFP/DsRed fractions regardless of MG132 treatment, and thus the $\log_2$(Cy5/Cy3) ratio of the probe corresponding to that ORF should be close to 0 (Cy5 ≈ Cy3) on all chips. In contrast, because cells expressing a specific EGFP-ORF that disappears from one cell fraction should reappear in other fractions, probes for proteasome substrates should show a negative value of $\log_2$(Cy5/Cy3) ratio on chips representing low EGFP/DsRed ratio cells and a positive value on chips of high EGFP/DsRed ratio cells. Combining data from the seven hybridizations improved the accuracy of discriminating true-positives from false-positives caused by experimental artifacts. Representative results for proteasome substrates and nonsubstrates are shown in Fig. 5C, and the complete screen results are in table S6. Consistent with the fact that proteasome degradation is the main pathway for cellular protein turnover (*29*), we found that the stability of >80% of EGFP-ORF proteins increased in cells treated with MG132, and none showed a decrease.

We chose a random sample of 85 ORFs predicted to encode proteasome substrates and 6 ORFs that do not and individually measured the EGFP/DsRed ratios of DsRed-IRES-EGFP reporter cell lines expressing those ORFs in the presence or absence of MG132. In all cases examined, the EGFP/DsRed measurements were in agreement with the array results (Fig. 5D). We detected great similarities between the EGFP/DsRed profiles inferred from arrays and those obtained by FACS measurements (Fig. 5D). To demonstrate that stabilization by MG132 was independent of the EGFP fusion, we C-terminally HA-tagged 24 ORFs and determined the steady-state concentration of their encoded proteins by Western blotting in the presence or absence of MG132. The abundance increased with MG132 treatment in all cases (Fig. 5E). These data collectively support the use of comparative GPS profiling as a general and powerful platform to identify specific protein turnover events in response to environmental changes.

Protein turnover represents an underexamined dimension of proteomics. We established a novel assay as a read-out for protein half-life, GPS, and combined it with DNA microarray deconvolution to allow highly parallel multiplex analysis of protein stability of over 8000 human proteins. Protein stabilities derived from GPS profiling are consistent with those reported in the literature and respond appropriately to different genetic backgrounds. Furthermore, unbiased validations for both the global protein stability screen and the comparative protein stability profile for proteasome substrates strongly supported the validity of the GPS analysis. There are, however, limitations to GPS. In certain subsets of proteins, such as membrane proteins and mitochondrial proteins, the presence of an N-terminal EGFP tag may alter their localization and hence their turnover. This can be addressed by using a C-terminal EGFP-tagged library. Because EGFP-ORF and DsRed proteins are translated through different

mechanisms, it is possible that the product of a specific ORF affects IRES function and, thus, influences the EGFP/DsRed readout in that particular cell. There will also be cases where the EGFP fusion may affect stability by affecting folding or obscuring degron sequences. Despite these limitations, this method offers a very deep window into a critical aspect of cellular physiology. The only other comprehensive global analysis of protein turnover that has been available was performed in budding yeast by using >3800 individual cycloheximide-chase analyses (*28*), a method that is impractical for mammalian cells.

A number of general findings emerged from this analysis. We found a bimodal distribution of protein half-lives centered around 0.5 and 2 hours. A similar distribution was previously observed in yeast, although with a shorter scale (*28*), which may be explained by the shorter cell cycle of yeast (~2 hours) compared with mammalian cells (~20 hours). We also find that longer proteins are relatively more stable. One possible explanation is that cells require more resources to synthesize longer proteins and tend to protect their investment. Although PEST sequences (polypeptide sequences enriched in proline, glutamic acid, serine, and threonine) are widely thought to be associated with short–half-life proteins (*30*), we found no enrichment of PEST sequences in labile proteins. Instead, unstable proteins appear to be rich in amino acids that can be phosphorylated, such as tyrosine and threonine. Indeed, phosphorylation is frequently a signal for regulated protein degradation (*31*). Thus, we conclude that the PEST hypothesis is incorrect in a general sense.

Proteins with unstructured regions (UPRs) are susceptible to degradation by the 20*S* proteasome in vitro (*32*). However, we found no correlation between the presence of UPRs (in both length and number) and protein instability. Because many UPRs function in molecular recognition, it is possible that in vivo UPRs are no longer "unstructured" and are protected by binding to their biological targets (*33*). Similar observations were made in yeast (*34*).

This GPS technology has a number of applications. It could be used to identify mutations that affect basal protein stability, which would reveal degron or stabilization sequences. GPS profiling could also be used to identify proteins whose stabilities change in response to stimuli, as well as during developmental transitions. GPS can be used to discover ubiquitin ligases or other proteins that regulate the stability of a protein of interest by coupling GPS with loss-of-function (from RNA interference) or gain-of-function screens that alter the DsRed/EGFP ratio. Conversely, this method could be used to identify substrates of ubiquitin ligases, currently a very labor-intensive endeavor with few general solutions, as we have done with the Skp1–cullin–F-box (SCF) ubiquitin ligase (*35*). GPS could be coupled with chemical screens to search for compounds that destabilize a protein of interest as opposed to inhibiting its activity by direct binding. GPS profiling could also be used to generate disease-specific protein stability signatures that may be useful for both diagnosis and elucidation of disease mechanisms. Finally, the integration of global protein stability information with other data sets will provide a global vision of regulatory networks with greater clarity and will help identify cross-talk between protein turnover and other levels of biological regulation (*36*, *37*). Thus, GPS has opened many avenues for protein-turnover studies.

References and Notes

1. S. Ghaemmaghami *et al.*, *Nature* **425**, 737 (2003).
2. S. F. Kingsmore, *Nat. Rev. Drug Discov.* **5**, 310 (2006).
3. J. M. Pratt *et al.*, *Mol. Cell. Proteomics* **1**, 579 (2002).
4. R. W. King, R. J. Deshaies, J. M. Peters, M. W. Kirschner, *Science* **274**, 1652 (1996).
5. V. Jesenberger, S. Jentsch, *Nat. Rev. Mol. Cell Biol.* **3**, 112 (2002).
6. A. Hershko, A. Ciechanover, *Annu. Rev. Biochem.* **67**, 425 (1998).
7. K. I. Nakayama, K. Nakayama, *Nat. Rev. Cancer* **6**, 369 (2006).
8. A. Ciechanover, A. L. Schwartz, *Biochim. Biophys. Acta* **1695**, 3 (2004).
9. A. L. Schwartz, A. Ciechanover, *Annu. Rev. Med.* **50**, 57 (1999).
10. L. Banks, D. Pim, M. Thomas, *Trends Biochem. Sci.* **28**, 452 (2003).
11. G. Nalepa, M. Rolfe, J. W. Harper, *Nat. Rev. Drug Discov.* **5**, 596 (2006).
12. Y. Sun, *Cancer Biol. Ther.* **2**, 623 (2003).
13. J. R. Newman *et al.*, *Nature* **441**, 840 (2006).
14. A. Sigal *et al.*, *Nature* **444**, 643 (2006).
15. A. Sigal *et al.*, *Nat. Methods* **3**, 525 (2006).
16. Materials and methods are available as supporting material on *Science* Online.
17. S. M. Ngoi, A. C. Chien, C. G. Lee, *Curr. Gene Ther.* **4**, 15 (2004).
18. T. V. Pestova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7029 (2001).
19. S. Matsuzawa, M. Cuddy, T. Fukushima, J. C. Reed, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14982 (2005).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. M. Welcker *et al.*, *Mol. Cell* **12**, 381 (2003).
22. X. Ye *et al.*, *J. Biol. Chem.* **279**, 50110 (2004).
23. J. Jin *et al.*, *Genes Dev.* **17**, 3062 (2003).
24. D. M. Koepp *et al.*, *Science* **294**, 173 (2001).
25. M. Scheffner, J. M. Huibregtse, R. D. Vierstra, P. M. Howley, *Cell* **75**, 495 (1993).
26. D. Lane, E. Harlow, *Nature* **298**, 517 (1982).
27. J. F. Rual *et al.*, *Genome Res.* **14**, 2128 (2004).
28. A. Belle, A. Tanay, L. Bitincka, R. Shamir, E. K. O'Shea, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13004 (2006).
29. K. L. Rock *et al.*, *Cell* **78**, 761 (1994).
30. M. Rechsteiner, S. W. Rogers, *Trends Biochem. Sci.* **21**, 267 (1996).
31. T. Hunter, *Mol. Cell* **28**, 730 (2007).
32. S. Prakash, L. Tian, K. S. Ratliff, R. E. Lehotzky, A. Matouschek, *Nat. Struct. Mol. Biol.* **11**, 830 (2004).
33. G. Asher, N. Reuven, Y. Shaul, *Bioessays* **28**, 844 (2006).
34. P. Tompa, J. Prilusky, I. Silman, J. L. Sussman, *Proteins* **71**, 903 (2008).
35. H.-C. S. Yen, S. J. Elledge, *Science* **322**, 923 (2008).
36. T. Ideker *et al.*, *Science* **292**, 929 (2001).
37. T. J. Griffin *et al.*, *Mol. Cell. Proteomics* **1**, 323 (2002).

38. We thank M. Vidal, J.W. Harper, G. Hu, and J. Jin for the hORFome v1.1 library; J. Daly and S. Lazo-Kallanian for FACS assistance; and J. Love, S. Gupta, and K. Hurov for helpful discussions. H.-C.S.Y. is a Jane Coffin Childs Memorial Fund fellow. D.M.C. is supported by the National Science Foundation. This work is supported by an NIH grant AG11085 to J.W.H. and S.J.E. S.J.E. is an investigator with the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/918/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S6

14 May 2008; accepted 22 September 2008
10.1126/science.1160489

Identification of SCF Ubiquitin Ligase Substrates by Global Protein Stability Profiling

Hsueh-Chi Sherry Yen and Stephen J. Elledge*

Ubiquitin-mediated proteolysis regulates all aspects of cellular function, and defects in this process are associated with human diseases. The limited number of identified ubiquitin ligase–substrate pairs is a major bottleneck in the ubiquitin field. We established and applied genetic technologies that combine global protein stability (GPS) profiling and genetic perturbation of E3 activity to screen for substrates of the Skp1–cullin–F-box (SCF) ubiquitin ligase in mammalian cells. Among the >350 potential substrates identified, we found most known SCF targets and many previously unknown substrates involved in cell cycle, apoptosis, and signaling pathways. Exploring cell cycle–stage stability, we found that several substrates used the SCF and other E3s in different cell cycle stages. Our results demonstrate the potential of these technologies as general platforms for the global discovery of E3-substrate regulatory networks.

The selective degradation of proteins is an important means of regulating gene expression and has pivotal roles in the control of various cellular processes. Ubiquitin (Ub)–mediated proteolysis is the major nonlysosomal proteolytic pathway in the cell and is required for the degradation of key regulatory proteins that include tumor suppressors, oncoproteins, and cell cycle regulators. Most proteins degraded by this pathway are first tagged with polyubiquitin chains by

the sequential action of three classes of enzymes: E1 (Ub-activating enzyme), E2 (Ub-conjugating enzyme), and E3 (Ub ligase). Once ubiquitinated, proteins are rapidly hydrolyzed by the 26*S* proteasome. Substrate specificity is largely conferred by E3 Ub ligases, and the interaction of substrates with E3s is the major point of regulation in Ub-mediated protein turnover (*1*).

There are more than 500 E3s in the human genome, but functional information is available for only a small fraction. Linking an E3 with its substrates is difficult and is generally dependent on either a functional connection or a physical association between the proteins. Given the large number of ubiquitinated substrates and E3s, more efficient strategies to deduce E3-substrate pairs are needed. Performing biochemical screens for E3 substrates is labor-intensive and is hampered by low substrate levels and intrinsically weak interactions between E3s and their substrates. Many E3s participate in human disease (*2*, *3*). Thus, elucidating E3-substrate interactions is not only critical to further our understanding of normal protein turnover control and its deregulation in disease, but may also provide valuable information for the development of new therapeutic intervention strategies.

The SCF ubiquitin ligase is a modular RING-type E3 and consists of at least four components: Skp1, Cul1, Rbx1, and an F-box protein. Cul1 functions as a scaffold that simultaneously interacts with Rbx1, coupled to an E2 enzyme, and with Skp1 to recruit an F-box protein, the subunit responsible for substrate recognition (*4*–*6*). About 70 F-box proteins have been discovered in humans, and several have been shown to be involved in diseases, such as Skp2 and Fbw7 in cancer (*7*, *8*). Many SCF substrates have been identified and are involved in a broad range of cellular functions that include cell cycle progression (e.g., cyclins and Cdc25A); signal transduction (e.g., Notch and inhibitor of nuclear factor κB); and transcription (e.g., SMAD and c-Jun) (*7*, *9*).

Here, we present the development of high-throughput approaches to provide a general solution for E3 substrate identification. We used global protein stability (GPS) profiling (*10*), coupled with genetic ablation of E3 function, to isolate new SCF substrates (*11*). We demonstrated the feasibility of these approaches through confirmation of known substrates and the discovery of a large number of new substrates. In addition to their use in E3 target identification, these strategies can be further generalized to detect proteins whose stabilities increase or decrease in response to various drugs or stimuli.

Inhibition of SCF activity by dominant-negative Cul1. To search for substrates of the SCF in vivo, we developed methods to quickly ablate cellular Cul1 function. Prolonged SCF inhibition has the potential to alter cell cycle distribution and to confound our analysis. Because Cul1 is very stable, we applied a dominant-negative (DNCul1), instead of small interfering RNA, to reduce Cul1 activity. To monitor SCF activity, we established reporter human embryonic kidney HEK 293T cell lines expressing enhanced green fluorescent protein (EGFP) fused to Cdc25A or $p21^{CIP1}$, two known SCF substrates, from the GPS reporter cassette (*10*). In these cells, *Discosoma* sp. red fluorescent protein (DsRed) and EGFP-fusions are expressed under the control of the same promoter, and thus the EGFP/DsRed ratio reflects

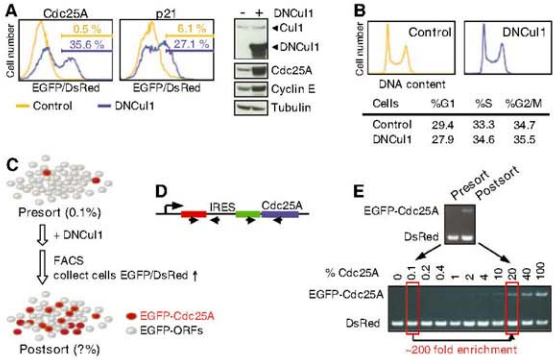


Fig. 1. A proof-of-principle screen for SCF substrates. **(A)** HEK 293T cells expressing DsRed-IRES-EGFP-Cdc25A or DsRed-IRES-EGFP-p21 were infected with either control or DNCul1 viruses for 18 hours and the EGFP/DsRed ratios of cells were measured by FACS. The abundance of Cul1 and endogenous SCF targets was analyzed by Western blot. **(B)** The cell cycle profile of HEK 293T cells infected with control or DNCul1 viruses for 18 hours was analyzed by propidium iodide staining. **(C)** Schematic illustration of the proof-of-principle screen. **(D)** Arrows represent PCR primers for DsRed and EGFP-Cdc25A. **(E)** Genomic DNA from presorted, postsorted, and various EGFP-Cdc25A and library cell mixtures was subjected to PCR and analyzed by agarose gel electrophoresis.

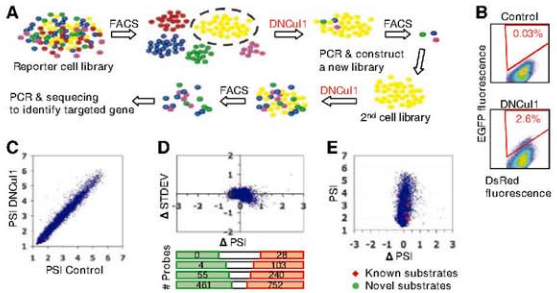


Fig. 2. Isolation of SCF targets by GPS profiling. **(A)** Schematic for the enrichment approach to E3 substrate identification. Cells with different EGFP/DsRed ratios are labeled with different colors. Shown here are the procedures for two rounds of enrichment. **(B)** The EGFP and DsRed fluorescence of a secondary cell library infected with control or DNCul1 viruses. **(C)** The PSI of EGFP-ORFs in control or DNCul1 virus–infected cells. **(D)** Changes in PSI and SD (STDEV) (Δ = DNCul1 minus control) were plotted, and the number of probes matching various ranges of ΔPSI is shown on the bottom of the graph. Note that some ORFs have two probes spotted on each microarray. **(E)** A plot of the PSI versus ΔPSI. Results from probes against previously discovered SCF targets are colored red, and those from probes against ICAT, APBB1IP, and FBXL14 are marked in green.

Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: selledge@genetics.med.harvard.edu

the stability of the EGFP-fusion protein and is not affected by changes in transcription (*10*). Blocking SCF activity should result in accumulation of EGFP-Cdc25A and EGFP-p21 and should increase EGFP/DsRed ratios. We compared several delivery conditions and found that lentiviral delivery of DNCul1 provided the most potent inhibition. As soon as 18 hours after viral infection, ~30% of reporter cells displayed an increased EGFP/DsRed ratio, and endogenous SCF substrates accumulated as well (Fig. 1A). Note that because the degradation of Cdc25A and p21 depends on cell cycle stage–specific phosphorylation, not all cells shift their EGFP/DsRed ratio in asynchronous cultures. The amount of DNCul1 sufficient for substrate stabilization under this condition did not alter the cell cycle profile of these cells, which made it possible to search for SCF targets without the confounding effects of cell cycle perturbation (Fig. 1B).

A proof-of-principle screen. We constructed a HEK 293T reporter cell library expressing EGFP fused to ~8000 human open reading frames (ORFs) from the GPS reporter cassette in their genome (*10*). Because the EGFP/DsRed ratio of these cells reflects the stability of the EGFP-fusion protein, SCF targets can be detected in cells that show increased EGFP/DsRed ratios in response to Cul1 inhibition. To test our ability to identify SCF substrates, we added cells carrying EGFP-Cdc25A into the EGFP-ORF cell library to see whether those EGFP-Cdc25A cells were recovered with our screen (Fig. 1C). EGFP-Cdc25A–expressing cells were mixed with EGFP-ORF library cells at a ratio of 1 to 1000 and infected with DNCul1 viruses. Cells with elevated EGFP/DsRed ratios were collected by fluorescence-activated cell sorting (FACS). To distinguish Cdc25A from library cells, genomic DNA was isolated and subjected to polymerase chain reaction (PCR) with two pairs of primers, one for DsRed and the other for EGFP-Cdc25A (Fig. 1D). The percentage of EGFP-Cdc25A cells in a mixed population was estimated by the ratio of PCR-amplified EGFP-Cdc25A to DsRed (Fig. 1E, bottom). The proportion of EGFP-Cdc25A cells in the FACS-isolated population increased to ~20%, a 200-fold enrichment compared with the starting mixture (Fig. 1E). These results support the idea that an enrichment strategy can identify SCF substrates.

Screening for SCF substrates by GPS profiling. We used two methods for identifying SCF substrates using GPS profiling with conditional Cul1 inactivation. The first method entailed presorting the reporter cell library into sublibraries with narrow, but defined, ranges of EGFP/DsRed ratios by FACS. Within each individual subpopulation, cells that showed increased EGFP/DsRed ratios in response to Cul1 inhibition were collected, and ORFs were recovered by PCR. Isolated ORFs were either sequenced or used to construct a secondary cell library for further enrichment (Fig. 2A). We applied two-rounds of selection with one sublibrary and identified enriched clones as predicted by our Cdc25A enrichment test (Fig. 2B and fig. S1). By sequencing 48 randomly chosen clones, we obtained 11 unique genes and confirmed that four of these encode proteins degraded by SCF, including $p21^{CIP1}$ and three novel substrates, ICAT (inhibitor of β-catenin and Tcf-4), APBB1IP (β-amyloid precursor protein–binding, family B, member 1–interacting protein), and FBXL14 (F-box and leucine-rich–repeat protein 14) (Figs. 3 and 4). The recovery of both known and new targets indicates that enrichment is a feasible method for substrate identification.

The second and more powerful approach used GPS profiling coupled with microarray deconvolution (*10*). In brief, the library cells infected with control or DNCul1 lentiviruses were fractionated into seven pools on the basis of their EGFP/DsRed ratios. The ORF sequences acting as unique identifiers for reporter cells were isolated from the genomic DNA from each pool by PCR amplification and quantified by microarray hybridization (fig. S2). The protein stability index (PSI) and standard deviation (SD), representing stability deviation from the mean, of each EGFP-ORF was calculated from the hybridization signals (*10*). SCF targets were identified by comparing the PSI of ORFs expressed in cells with and without DNCul1 virus infection.

We performed a series of analyses and confirmed that the quality of our microarray hybridization is high (*11*) (table S1). We plotted the PSI of each EGFP-ORF protein from control and DNCul1-expressing cells and observed a high correlation coefficient of 0.982 with a slope of nearly 1 (1.079) (Fig. 2C). The high degree of identity between PSIs from cells with and without DNCul1 supports the reliability of the array approach for measuring protein stability and confirms that most EGFP-ORF proteins are not tar-

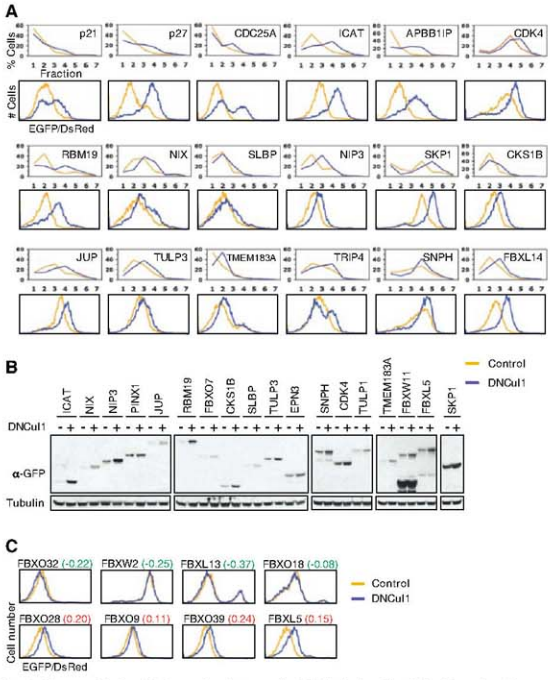


Fig. 3. Screen validation. **(A)** Comparison between the EGFP/DsRed profiles derived from the microarray data (top) and from individual FACS analyses (bottom). **(B)** The EGFP-ORF protein level in cells expressing the indicated GFP fusion protein infected with control (–) or DNCul1 (+) viruses. **(C)** The EGFP/DsRed ratios of cells infected with control or DNCul1 viruses were analyzed by FACS. The number shown at the top of each graph is the ΔPSI of the indicated F-box protein from the screen.

geted by the SCF. We calculated the difference in PSI with and without DNCul1 (Δ = DNCul1 minus control) and found more ORFs with increased stability than decreased stability after SCF inhibition (Fig. 2D). Because inhibiting the SCF should lead to stabilization, the detection of more ORFs with a positive ΔPSI supports the validity of this approach. Because noise from low hybridization signals can be a problem for microarray-based detection, we examined the relation between ΔPSI and hybridization signal. Low signals were not the reason for the observed stability variation (fig. S3).

To further assess the validity of this strategy, we examined the behavior of known SCF substrates in the screen. We detected a significant ΔPSI for 73% (11 of 15) of previously described SCF substrates present in the library (Fig. 2E and table S2). Moreover, all three novel substrates isolated from our enrichment screen, ICAT, APBB1IP, and FBXL14, were also recovered. These analyses collectively reveal that our experimental approach is capable of global isolation of E3 substrates.

SCF substrate validation. We selected ORFs for subsequent validation studies by the following criteria. First, to reduce false-positives generated from spurious hybridization, we stipulated that the Cy5 channel signal for a given probe had to be greater than fivefold above background, and the variation in Cy5 intensity of the same probe between chips must be less than threefold. Second, the PSI had to increase more than 0.25 units when SCF was inhibited. The selection of this threshold was referenced to the ΔPSI of known SCF substrates. Third, we focused on proteins that are normally unstable, that is, the PSI of those proteins from control cells must be less than 3.3. Last, for ORFs with two probes, results from both probes had to be consistent. As a result, 359 ORFs met these criteria, and we chose 66 ORFs independent of their known functions for validation (table S3).

As an initial validation of the hits from our screen, ORFs encoding potential SCF substrates were individually recombined into the GPS reporter to create stable cell lines. Each reporter cell line was infected with either control or DNCul1 viruses, and their EGFP/DsRed ratios were measured by FACS. Clones with increased EGFP/DsRed in response to DNCul1 expression were independently tested at least three times. Of the 66 tested, we confirmed 31 lines whose EGFP/DsRed ratio was dependent on the SCF (Table 1). We next compared the EGFP/DsRed profile derived from the microarray data with those from individual FACS analysis. The EGFP/DsRed pattern from bulk analysis with the microarray was similar to that from single-tube FACS (Fig. 3A), which suggested that the isolation of those ORFs was not due to spurious factors. The advantage of this microarray-based method is that it identifies targets and also precisely reports the stability of a protein and its degree of change after SCF inhibition. We further confirmed protein abundance by Western blot analysis of cell extracts with antibodies to EGFP (Fig. 3B). The validation rate suggests that the signal-to-noise ratio of this screen is high and demonstrates the robustness of this approach as a general method to identify E3 substrates.

Validated proteins include 6 previously reported and 25 SCF substrates not previously described that function in various cellular activities, including cell cycle progression, apoptosis, and signaling (Table 1). We also identified Skp1 and five F-box proteins that constitute the SCF itself. Because many E3 ligases control their own turnover, the isolation of SCF subunits suggests that their degradation might be mediated through autoubiquitination within the complex. In fact, two budding yeast F-box proteins, Cdc4p and Grr1p, have been shown to ubiquitinate themselves (*12*, *13*). We therefore wondered whether the dependence on the SCF for turnover is a general feature for F-box proteins. After surveying our microarray data for every F-box protein in the library, we found that not all F-box proteins were stabilized after SCF inactivation (table S4). To explore this further, we randomly selected four F-box proteins that did not increase PSI after DNCul1 treatment and four that mildly increased but did not meet the criteria for validation (ΔPSI > 0.25) and tested them by FACS. The results were consistent with those from the screen (Fig. 3C). Thus, degradation by the SCF is not a universal feature of all F-box proteins. The recovery of a large number of previously described substrates together with the high validation rate supports the idea that many proteins on the candidate substrate list are likely to be bona fide targets of the SCF.

To further examine the validity of this method, we tested whether the observed degradation by SCF depends on the N-terminal EGFP fusion. We analyzed the amount of either endogenous proteins (Fig. 4A) or proteins tagged with a single copy of hemagglutinin (HA) at the C terminus (Fig. 4B) in cells with and without DNCul1. In most cases (89%, 17 of 19), proteins accumulated to greater steady-state amounts after loss of SCF activity (Table 1). Because ORF-HA proteins were expressed under the control of the ubiquitously active elongation factor EF1α promoter whose activity is not affected by the SCF, the detected increase in protein abundance is likely due to an increase in protein stability and not synthesis. To confirm it, we measured the half-life of proteins in cells in which protein synthesis had been inhibited with cycloheximide. All tested proteins were stabilized in cells infected with DNCul1 viruses (Fig. 4C), which suggested that the EGFP/DsRed ratio serves as a reliable indicator of protein stability. To test whether the SCF substrates isolated from HEK 293T cells are also targeted by the SCF in other cell types, we analyzed the amounts of endogenous proteins for which antibodies are available (Fig. 4D). The proteins accumulated in response to DNCul1 in most cell types, but the degree of accumulation was cell-type specific, which suggested that various cell types may have distinct combinations of proteolysis pathways (for example, differential expression profiles of F-box proteins).

SCF substrates and the cell cycle. A unique feature of GPS is its ability to monitor protein turnover at the resolution of single living cells. Measuring the average of a cell population can overlook events that happen in only a subset of cells, all-or-none effects, and variability between cells. We combined GPS and Hoechst staining for DNA content to study cell cycle–mediated protein degradation by three-color FACS analysis. We expressed several cell cycle proteins from the GPS cassette and found that cells that carry

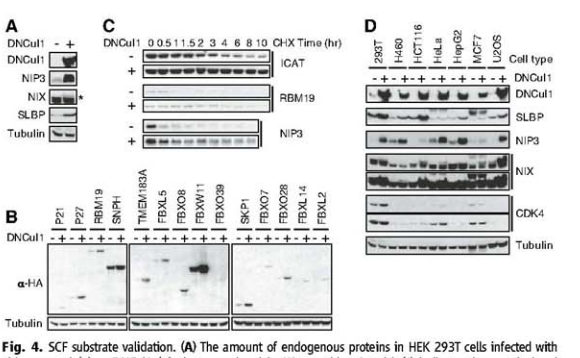


Fig. 4. SCF substrate validation. **(A)** The amount of endogenous proteins in HEK 293T cells infected with either control (–) or DNCul1 (+) viruses analyzed by Western blot. Asterisk (*) indicates the protein band representing NIX according to its molecular weight. **(B)** Western blot analysis of the indicated HA-tagged ORFs from cells infected with either control (–) or DNCul1 (+) viruses. HA-ORFs were expressed under the control of the EF1α promoter and introduced into HEK 293T cells by retroviral infection at low MOI. **(C)** Protein stability measured by cycloheximide (CHX)-chase. **(D)** The level of endogenous proteins in various cell lines infected with control (–) or DNCul1 (+) viruses. Results from two exposure times are shown for NIX and CDK4.

EGFP-Cdc20, -cyclin A, -cyclin E, and -securin fusion proteins displayed more divergent EGFP/DsRed ratios than those of the control EGFP-degron cell series (*10*) (Fig. 5A). We found that cells expressing EGFP-Cdc20, but not the control, d4EGFP (mutant of a degron with $t_{1/2}$ of 4 hours), show distinct EGFP/DsRed ratios in cells with different DNA contents (Fig. 5B), which indicates that the variability in EGFP/DsRed ratios is due to differences in stabilities in different cell cycle stages. Consistent with the fact that Cdc20 is degraded by the anaphase-promoting complex/cyclosome APC/C^{CDH1} during late mitosis (*14*, *15*), two populations of EGFP/DsRed ratios were observed from EGFP-Cdc20–expressing cells with G_2/M DNA content, with the lowest observed EGFP/DsRed ratio in G_1 cells. Thus, GPS combined with Hoechst staining provides a way to rapidly identify proteins whose turnovers are cell cycle–regulated.

Many proteins with cell cycle–regulated stabilities are controlled by the SCF, so we wondered if some SCF substrates identified from our screen might also be cell cycle–regulated. Indeed, the half-lives of p27, stem-loop binding protein (SLBP), and synaphilin (SNPH) varied during cell cycle progression (Fig. 5C and fig. S4A). The stabilization of p27 in G_1 and SLBP during S phase is consistent with previous reports (*16*, *17*), and their known cell cycle stage–specific functions. We found SNPH, a protein involved in mitochondrial docking and vesicular transport (*18*), is unstable specifically in G_1. We next asked whether the degradation of these proteins by the SCF is restricted to a particular cell cycle stage (Fig. 5D and fig. S4B). For some proteins, the stabilization by DNCul1 was more pronounced in, but not restricted to, specific cell cycle stages, such as p27 and tubby like protein 3 (TULP3) during S and G_2 phases and Bcl2/adenovirus E1B 19-kD–interacting protein 3 (NIP3) and Cdc28 protein kinase regulatory subunit 1B (CKS1B) during S phase (Fig. 5D). In contrast, degradation of SLBP and SNPH by the SCF is specifically limited to G_2/M and G_1 phase cells, respectively. Although SLBP was unstable in both the G_1 and G_2/M phases (Fig. 5C), stabilization of SLBP by DNCul1 was detected only at G_2/M (Fig. 5D), which suggested that there are at least two distinct ligases or pathways regulating SLBP turnover: the SCF in G_2/M and an unknown ligase in G_1.

The phosphorylation of threonines 61 and 62 is required for SLBP degradation but the E3 ligase remains unidentified (*16*). We mutated the two threonines to alanines and found that the mutations led to G_2/M phase-specific stabilization of SLBP (Fig. 5E). Furthermore, the half-life of mutant SLBP was no longer regulated by the SCF. The EGFP/DsRed profile of cells expressing mutated SLBP with or without DNCul1 is the same as that of cells carrying wild-type SLBP with DNCul1 at all cell cycle stages, which suggests that threonines 61 and 62 are essential components of the degron responsible for SCF-mediated degradation.

The assignment of substrates to their cognate E3 enzymes has proven difficult and has primarily relied on physical association (*19*–*21*) or the presence of known degron motifs (*22*, *23*), or has depended on genetic screens in model organisms (*24*–*28*). Each of these methods has significant drawbacks and, so far, has identified only a few dozen substrates in total. A recent study in yeast suggested that in vivo screens may serve as a better alternative for identifying physiologically relevant E3 substrates (*29*). They isolated SCF^{Grr1} substrates by using the budding yeast ORF-GFP fusion strain collections mated to *grr1* mutants and sporulated in a 96-well format followed by quantitative microscopy. As the GFP fusions used were under control of their endogenous promoters, the results were confounded to some degree by transcriptional control. Although the method was successful, it is, unfortunately, not applicable to mammals.

In this study, we employed GPS profiling coupled with either an enrichment strategy or a microarray deconvolution approach to search for SCF substrates in mammalian cells. Using a library of ~8000 ORFs, we successfully isolated both previously reported and novel SCF substrates, which demonstrated the potential of these technologies as systematic platforms for high-throughput discovery of E3 substrates. Our results from the microarray-based screen are particularly encouraging. We recovered 73% of previously described SCF substrates in our library and generated a list of likely substrates containing 359 proteins. Individual FACS measurements confirmed 31 of the 66 tested proteins as bona fide substrates. Furthermore, results from ORFs with only a slight increase in stability were readily reproducible. The low false-positive and false-negative rates of the screen are favorable compared with those of most high-throughput microarray-based approaches. Approximately 300 candidate substrates from the screen remain to be characterized, and we conservatively estimate that as many as 100 SCF substrates or more may be identified by this screen. This is an underestimate of the total number of SCF substrates because the current library is not

Table 1. Summary of validation results. The asterisk (*) marks the ORF-HA proteins whose expression was not detected by Western blot. Because ORF-HA was expressed as a single copy from the genome, it is likely that a single HA epitope tag is insufficient for the detection of low-abundance proteins. Proteins that were not tested are labeled N/A (not assessed). Except APBB1IP, all ORFs isolated from the screen are full-length. ID, identification.

Protein name	Gene ID	ΔPSI	Known?	FACS	Western blot	
					Anti-GFP	Anti-HA or endogenous
APBB1IP/PREL1	54518	1.22	No	+	+	N/A*
ICAT/CTNNBIP1	56998	0.72	No	+	+	N/A*
RBM19	9904	0.67	No	+	+	+
NIX/BNIP3L	665	0.46	No	+	+	+
SLBP	7884	0.28	No	+	+	+
ARMCX6	54470	0.33	No	+	N/A	N/A*
PINX1	54984	0.55	No	+	+	–
EPN3	55040	0.57	No	+	+	–
TMEM183A	92703	0.26	No	+	+	+
NIP3/BNIP3	664	0.33	No	+	+	+
CKS1B	1163	0.31	No	+	+	N/A*
CDK2AP1	8099	0.32	No	+	N/A	N/A*
AASDHPPT	60496	0.27	No	+	N/A	+
FBXL14	144699	0.43	No	+	+	+
FBXL2	25827	0.43	No	+	N/A	+
TRIP4	9325	0.36	No	+	N/A	+
SKP1	6500	0.72	No	+	+	+
FBXO7	25793	0.26	No	+	+	+
SNPH	9751	0.42	No	+	+	+
PYROXD1	79912	0.27	No	+	N/A	+
FBXW11	23291	0.32	No	+	+	+
CDK4	1019	0.45	No	+	+	+
JUP	3728	0.41	No	+	+	N/A
TULP1	7287	0.36	No	+	+	N/A
TULP3	7289	0.45	No	+	+	N/A
p21	1026	0.37	Yes	+	+	+
p27	1027	0.64	Yes	+	+	+
CDC25A	993	0.29	Yes	+	+	+
NFKBIB	4793	0.32	Yes	+	N/A	N/A
USP18	11274	0.28	Yes	+	N/A	N/A
FBXO5	26271	0.27	Yes	+	N/A	N/A

complete, and the detection of some substrates may require an extended disruption of SCF activity. Moreover, the screen was performed in only a single cell type, which may not fully express the F-box proteins for some substrates or the signaling pathway components, such as kinases, needed to target proteins for degradation. These circumstances probably explain why some of the known SCF substrates thought to be present in the library were not identified in the screen. In addition, we cannot rule out the possibility that the EGFP fusion affects SCF-mediated degradation in a protein-specific manner.

We performed an analysis on the distribution of GO processes associated with the 359 candidate substrates and found that these proteins are involved in diverse cellular functions with a minor enrichment in apoptosis, which suggests that the SCF plays critical roles in a broad range of cellular activities and is not limited to cell cycle regulation. For example, ICAT is a negative regulator of the Wnt pathway. ICAT inhibits β-catenin nuclear signaling by binding to β-catenin and competing for its interaction with TCF transcription factors (*30*). The *ICAT* gene is located at a frequent target for LOH in many human cancers (*31*). Conversely, the F-box protein required for its degradation may act as an oncogene. NIX and NIP3 are functional homologs and members of the BCL2/adenovirus E1B–interacting protein family (*32*, *33*). NIX and NIP3 are transcriptionally induced during hypoxia, and their expression promotes apoptosis (*34*). Both NIX and NIP3 are degraded by the proteasome (*33*). SNPH regulates synaptic vesicle docking and fusion by inhibiting the SNARE (soluble *N*-ethylmaleimide–sensitive factor attachment protein receptor) complex and is required for proper distribution of mitochondria within axons on neurons (*18*, *35*). SNPH was unstable in G_1, and its degradation by the SCF is G_1-specific. The pivotal role of the SCF in cell cycle progression was further highlighted through our identification of cell cycle proteins as novel SCF substrates, such as CKS1B, a Cdk-binding protein that is also an accessory protein necessary for SCF^{Skp2} to target p27 degradation (*36*), and SLBP, the sole cell cycle–regulated factor required for histone mRNA processing (*16*, *37*). Overexpression of CKS1B has been observed in many tumors and correlates with poor prognosis (*38*). The stability of SLBP is regulated during the cell cycle, but the E3s controlling its degradation are not known (*16*). We found that there are distinct pathways mediating SLBP turnover at G_1 and G_2/M, and the degradation of SLBP by SCF is G_2/M-specific and requires the phosphorylation of threonines 61 and 62. We also found several substrates, such as CKS1B, NIP3 and TULP3, that did not show cell cycle–regulated stability but, nonetheless, showed cell cycle–stage differences in their SCF-dependent degradation. This suggests the existence of other E3s that work together with the SCF to control their degradation throughout the cell cycle.

Although inactivation of an E3 was our variable parameter, it is clear that the described system could be used to identify proteins whose stabilities change because of a number of different stimuli or genetic perturbations. Thus, one could envision looking for proteins whose stabilities change in response to chemical or physical stimuli such as DNA damage or hormones. Various screens could be performed with specific protein fusions from this expression cassette to identify pathways or chemicals that control its stability. The use of GPS profiling to measure perturbation of protein stability adds a new dimension to our ability to examine the proteome and provides a step toward the goal of a systems-level understanding of cellular physiology.

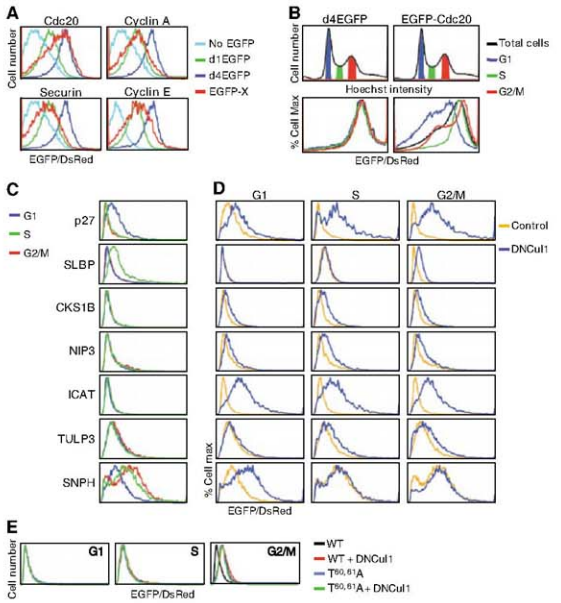


Fig. 5. SCF substrate stability and the cell cycle. **(A)** Overlay of the EGFP/DsRed ratios of EGFP-degron fusions (d1EGFP, $t_{1/2}$ = 1 hour; d4EGFP, $t_{1/2}$ = 4 hours) and EGFP fused to cell cycle proteins, as indicated in the graph. **(B)** Asynchronous cells expressing EGFP-Cdc20 fusions or the d4EGFP control from the GPS cassette were colored with Hoechst stain and analyzed by FACS. The EGFP/DsRed profiles of the asynchronous population or cells at a specific cell cycle stage, as judged by the DNA content, were compared. **(C)** Cells expressing EGFP-fused SCF substrates were analyzed as (B). **(D)** The EGFP/DsRed ratio of cells with control or DNCul1 viruses at different cell cycle stages. **(E)** The EGFP/DsRed ratio of cells carrying either wild-type or $Thr^{61,62}$Ala-mutated SLBP infected with control or DNCul1 viruses.

References and Notes

1. A. L. Schwartz, A. Ciechanover, *Annu. Rev. Med.* **50**, 57 (1999).
2. G. Nalepa, J. Wade Harper, *Cancer Treat. Rev.* **29** (Suppl. 1), 49 (2003).
3. L. T. Vassilev *et al.*, *Science* **303**, 844 (2004).
4. N. Zheng *et al.*, *Nature* **416**, 703 (2002).
5. C. Bai *et al.*, *Cell* **86**, 263 (1996).
6. D. Skowyra, K. L. Craig, M. Tyers, S. J. Elledge, J. W. Harper, *Cell* **91**, 209 (1997).
7. T. Cardozo, M. Pagano, *Nat. Rev. Mol. Cell Biol.* **5**, 739 (2004).
8. J. T. Winston, D. M. Koepp, C. Zhu, S. J. Elledge, J. W. Harper, *Curr. Biol.* **9**, 1180 (1999).
9. K. I. Nakayama, K. Nakayama, *Nat. Rev. Cancer* **6**, 369 (2006).
10. H.-C. S. Yen, Q. Xu, D. M. Chou, Z. Zhao, S. J. Elledge, *Science* **322**, 918 (2008).
11. Materials and methods are available as supporting material on *Science* Online.
12. P. Zhou, P. M. Howley, *Mol. Cell* **2**, 571 (1998).
13. J. M. Galan, M. Peter, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9124 (1999).
14. G. Fang, H. Yu, M. W. Kirschner, *Mol. Cell* **2**, 163 (1998).
15. S. Prinz, E. S. Hwang, R. Visintin, A. Amon, *Curr. Biol.* **8**, 750 (1998).
16. L. Zheng *et al.*, *Mol. Cell. Biol.* **23**, 1590 (2003).
17. A. C. Carrano, E. Eytan, A. Hershko, M. Pagano, *Nat. Cell Biol.* **1**, 193 (1999).
18. J. S. Kang *et al.*, *Cell* **132**, 137 (2008).
19. J. Jin, X. L. Ang, T. Shirogane, J. Wade Harper, *Methods Enzymol.* **399**, 287 (2005).

20. D. M. Koepp *et al.*, *Science* **294**, 173 (2001).
21. F. Margottin *et al.*, *Mol. Cell* **1**, 565 (1998).
22. C. M. Pfleger, M. W. Kirschner, *Genes Dev.* **14**, 655 (2000).
23. J. T. Winston *et al.*, *Genes Dev.* **13**, 270 (1999).
24. S. M. Siepka *et al.*, *Cell* **129**, 1011 (2007).
25. K. H. Moberg, D. W. Bell, D. C. Wahrer, D. A. Haber, I. K. Hariharan, *Nature* **413**, 311 (2001).
26. E. Spencer, J. Jiang, Z. J. Chen, *Genes Dev.* **13**, 284 (1999).
27. J. Jiang, G. Struhl, *Nature* **391**, 493 (1998).
28. D. C. Nelson, J. Lasswell, L. E. Rogg, M. A. Cohen, B. Bartel, *Cell* **101**, 331 (2000).
29. J. A. Benanti, S. K. Cheung, M. C. Brady, D. P. Toczyski, *Nat. Cell Biol.* **9**, 1184 (2007).
30. K. Tago *et al.*, *Genes Dev.* **14**, 1741 (2000).
31. J. Reifenberger *et al.*, *Int. J. Cancer* **100**, 549 (2002).
32. J. M. Boyd *et al.*, *Cell* **79**, 341 (1994).
33. G. Chen *et al.*, *J. Biol. Chem.* **274**, 7 (1999).
34. R. K. Bruick, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9082 (2000).
35. G. Lao *et al.*, *Neuron* **25**, 191 (2000).
36. D. Ganoth *et al.*, *Nat. Cell Biol.* **3**, 321 (2001).
37. Z. F. Wang, M. L. Whitfield, T. C. Ingledue 3rd, Z. Dominski, W. F. Marzluff, *Genes Dev.* **10**, 3028 (1996).
38. S. Kitajima *et al.*, *Am. J. Pathol.* **165**, 2147 (2004).
39. We thank M. Vidal, J. W. Harper, J. Jin, and G. Hu for the hORFeome v1.1 library; J. W. Harper for the DNCul1/pcDNA3 construct; D. M. Chou for experimental assistance and discussion; Q. Xu for bioinformatics analysis; M. Schlabach for microarray assistance; J. Daly and S. Lazo-Kallanian for their help with cytometry; and K. Hurov for suggestions. H.-C.S.Y. is a fellow of Jane Coffin Childs Memorial Fund. This work was funded by an NIH grant AG 11085 to J. W. Harper and S.J.E. S.J.E. is an investigator with the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/923/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S4

14 May 2008; accepted 22 September 2008
10.1126/science.1160462

Slow Electron Cooling in Colloidal Quantum Dots

Anshu Pandey and Philippe Guyot-Sionnest*

Hot electrons in semiconductors lose their energy very quickly (within picoseconds) to lattice vibrations. Slowing this energy loss could prove useful for more efficient photovoltaic or infrared devices. With their well-separated electronic states, quantum dots should display slow relaxation, but other mechanisms have made it difficult to observe. We report slow intraband relaxation (>1 nanosecond) in colloidal quantum dots. The small cadmium selenide (CdSe) dots, with an intraband energy separation of ~0.25 electron volts, are capped by an epitaxial zinc selenide (ZnSe) shell. The shell is terminated by a CdSe passivating layer to remove electron traps and is covered by ligands of low infrared absorbance (alkane thiols) at the intraband energy. We found that relaxation is markedly slowed with increasing ZnSe shell thickness.

When semiconductors are formed into quantum dots (QDs), discrete electronic states arise through confinement by the boundary. These states can be exploited in optical applications, where size-tunable narrow fluorescence emissions with lifetimes in the nanosecond range are useful (*1*). However, such narrow fluorescence emission entails the dissipation of any initially absorbed excess energy as heat within hundreds of femtoseconds (*2*, *3*). This rapid energy loss ("electron cooling") is not useful in electronics applications such as photovoltaics or infrared (IR) devices. Slower dissipation might give time to extract the energy of hot carriers for more efficient photovoltaics (*4*), and it would also enable the detection and emission of IR radiation via the intraband transitions of quantum dots.

In small QDs, such as colloidal CdSe, with a low electron mass, slow dissipation is in fact predicted on the basis that the lattice vibrations (30 meV) cannot couple widely separated electronic states (300 meV) (*5*–*8*) (Fig. 1B). The striking absence of such a "phonon bottleneck" in photoexcited QDs has been explained by fast excitonic cooling, where the electron transfers its energy to the much larger density of states of the more massive hole (*9*–*11*) (Fig. 1C). Nonetheless, efforts to decouple the electron and hole led to marginally longer picosecond relaxation times (*12*–*14*), raising the possibility of other mechanisms involving local intermediate states, generically called traps (*15*, *16*), and high-frequency local modes (*17*) or molecular vibrations (*14*).

In our experiments, electron cooling between the two lowest conduction band states of small colloidal CdSe dots, $1S_e$ and $1P_e$, is slowed to longer than 1 ns, which is more than three orders of magnitude slower than the relaxation time between the 1S and 1P exciton (*2*). This slower cooling is achieved by using a thick ZnSe shell to separate electrons and holes and to increase the distance of the electronic states from the ligands. We observed three complementary mechanisms shown schematically in Fig. 1. With an exposed ZnSe surface, electron trapping takes place on a time scale of 10 to 30 ps (Fig. 1D). Capping the ZnSe shell with one monolayer (1 ML) of CdS reduces electron trapping, but it can also prevent hole trapping and allows fast excitonic cooling (<6 ps) (Fig. 1C). A CdSe capping monolayer reduces electron trapping as well, but with hole-trapping ligands, electron cooling slows from 10 ps to >1 ns, in agreement with a mechanism of energy transfer to vibrations via dipole coupling (Fig. 1E).

In our experiments, the electron in the conduction band state of the dots is photogenerated, which leaves a hole initially in the valence band. Reducing electron-hole coupling requires a shell that preserves electron confinement and extracts the hole to a remote state, as shown in Fig. 2A. To avoid intermediate local electronic states or vibrational modes that could arise at defects, the shell must be epitaxial, and its surface should have no states that would trap the electron. To limit coupling to ligand vibrations, the shell should be thick and its outer surface capped by ligands of low IR absorbance.

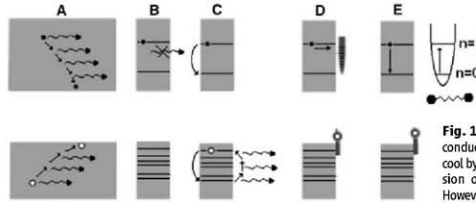


Fig. 1. (A) In bulk semiconductors, hot carriers cool by sequential emission of phonons. **(B)** However, in a small dot, electronic states are too far apart for phonon emission. **(C)** An electron can cool by transferring its energy to the hole with higher state density that then relaxes via phonons. **(D)** With a reduced coupling to the hole, shown here as surface-trapped, the electron can cool via intermediate trap states. **(E)** The electron can transfer its energy to a resonant high-frequency vibration.

James Franck Institute, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: pgs@uchicago.edu

To satisfy these criteria, we used a combination of shell materials resulting in a $CdSe/ZnS_{1ML}/ZnSe_{nML}/CdSe_{1ML}$ structure with the core CdSe in the zincblende structure (*18*). We used alkane thiolate ligands, which act as hole traps on chalcogenide QDs. They also have no hydrogen bonding mid-IR background, and their IR absorption is smallest between the CH stretches and CH bends, between 0.3 and 0.23 eV (*14*). This energy window determines where we set the $1S_e$-$1P_e$ intraband resonance and hence the core size. For the electron-confining/hole-extracting shell, we chose ZnSe because it has a weak type II alignment with CdSe, with a conduction band offset of ~1 eV (*19*). However, only thin ZnSe shells (<1 nm thick) have been grown previously (*20*). To grow a thick epitaxial and spherical ZnSe shell up to ~12 ML (each monolayer is ½ lattice constant, i.e., 0.28 nm), a high temperature (280°C) was required. To retain the monodispersed core size at that temperature, we found that an initial ZnS layer (~1 ML) was most helpful. Finally, with QDs terminated by ZnSe, the electrons were also easily trapped; this problem was solved by growing a final monolayer of CdSe.

The samples were characterized by optical spectroscopy and transmission electron microscopy (TEM) at all stages of the growth. The TEM images in Fig. 2B show the spherical and epitaxial shell growth. The standard deviation of the number of ZnSe monolayers (estimated from sizes) varies from <1 for thin shells to ~2 for the thickest shells. The optical spectra taken at various stages of growth are shown in Fig. 2C. With increasing ZnSe shell thickness, the weakly shifting lower-energy exciton peak (called the S-exciton) confirms the retention of monodispersed cores and the good electron-confining potential of ZnSe. The disappearance of several higher-energy features (e.g., the light hole to $1S_e$ transition seen at 510 nm in Fig. 2C) that arise from the discrete valence states is consistent with the increasingly delocalized hole in ZnSe (*19*). For the transient measurements, the cleaned QDs are dissolved in C_2Cl_4 and placed in a 1-mm-thick cell.

The electron cooling time was measured by a sequence of three optical pulses designed to (i) prepare the dots with one electron in $1S_e$, (ii) resonantly pump the electron from $1S_e$ to $1P_e$, and (iii) probe $1S_e$ recovery and $1P_e$ depopulation (*13*). The setup and procedure have been described for CdSe cores (*14*). To create a population of dots with one electron in $1S_e$, we use a green pulse (0.532 μm) above the S-exciton. Electrons and holes are created, and after multicarrier recombination (occurring within ~100 ps) and

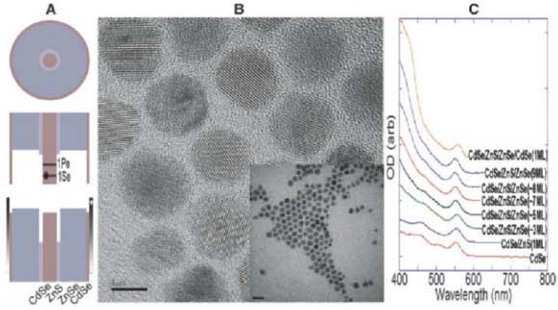
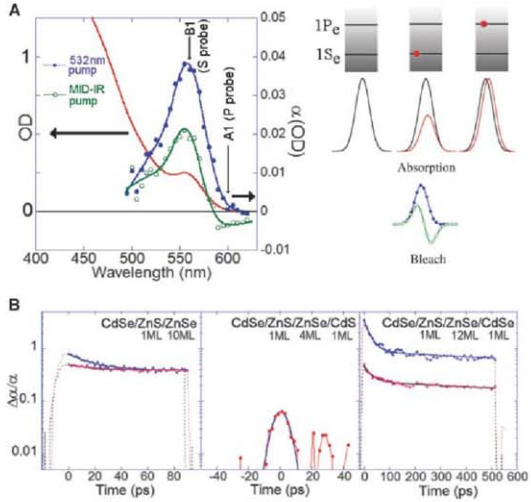


Fig. 2. **(A)** Schematic of the electron-confining and hole-extracting colloidal structure used to probe electron relaxation from $1P_e$ to $1S_e$. **(B)** TEM pictures of $CdSe/ZnS_{1ML}/ZnSe_{9ML}/CdSe_{1ML}$. Scale bar, 5 nm (inset, 20 nm). The CdSe core is ~3 nm in diameter. **(C)** Absorption spectra at various stages of shell growth.

Fig. 3. **(A)** Absorption spectrum (red line) and transient bleaches of a $CdSe/ZnS_{1ML}/ZnSe_{5ML}/CdSe$ sample taken 600 ps after the green pulse (blue circles) and 10 ps after the IR pump (green circles). The left schematic shows how a $1S_e$ electron mostly weakens the S-exciton absorption (from black to red), whereas a $1P_e$ electron mostly redshifts the S-exciton. The bleach after the green pump is due to state filling by the $1S_e$ electron. The bleach after the IR pump reflects state filling by $1S_e$ electrons and redshift caused by $1P_e$ electrons, which is most visible at A1. **(B)** Transient bleach $\Delta\alpha/\alpha$ at B1 (red dots) and A1 (blue dots) for the three different samples discussed in the text. The solid lines are biexponential fits (left and right graphs) and a Gaussian (center graph).



cooling, the dots are left with a single electron and a hole. The hole is not detectable in absorption, whether it is trapped or delocalized, because of weak overlap or high degeneracy, respectively. The electron in $1S_e$ can be clearly identified by the appearance of the strongly allowed $1S_e$-$1P_e$ intraband transition, as well as by the bleach of the S-exciton, because $1S_e$ is only doubly degenerate. Figure 3A shows a typical bleach probed by a tunable visible pulse 600 ps after the green pulse. The bleach has an optical density of 0.04, which corresponds to ~15% of the sample optical density and implies that ~30% of the dots have one electron in $1S_e$. The electron remains in $1S_e$ until radiative and nonradiative recombination with the hole or trapping. As such, recombination/trapping of the $1S_e$ electron takes place over ~1 ns for ZnSe-terminated surfaces and much longer (>>1 ns) with a CdS and CdSe outer monolayer.

At a varying delay after the green pump, an IR pump pulse (~4 μm, 20 μJ) resonantly excites the electron from $1S_e$ to $1P_e$. This excitation reduces the bleach in Fig. 3A to $\alpha - \Delta\alpha$, where α denotes the bleach without the IR pump. The probe for the $1S_e$ population is at the bleach peak, at position B1 in Fig. 3A. [The notation B1 was introduced in (*2*) for CdSe cores.] At B1, $\Delta\alpha$ mostly detects the change of the $1S_e$ occupation and $\Delta\alpha/\alpha$ gives the fraction of dots for which the IR pump removed the electron from $1S_e$. This is ~45% in Fig. 3A. The probe for the $1P_e$ population is at the position noted A1 (*2*) on the red edge where α is smaller. $\Delta\alpha$ at A1 is caused by both the reduced bleach that results from the loss of the $1S_e$ electron and the absorption by the redshifted exciton (*3*), as shown in the schematic in Fig. 3A. The redshift is explained by a stabilizing polarization interaction of the S-exciton with the electron charge in $1P_e$. The factor of 5 to 10 difference in the magnitude of α between A1 and B1 is normalized out by the ratio $\Delta\alpha/\alpha$. If the electron relaxes from $1P_e$ directly to $1S_e$, A1 and B1 follow identical kinetics but $\Delta\alpha/\alpha$ is larger at A1. If the electron leaves $1P_e$ by recombination/trapping, $\Delta\alpha$ reflects only the loss of the $1S_e$ electron; hence, it is a fractional change of α, and $\Delta\alpha/\alpha$ has identical values at the B1 and A1 positions.

Figure 3B shows the very different results obtained with different outer surfaces. With an outer ZnSe surface capped with amine and carboxylate ligands, $\Delta\alpha/\alpha$ at A1 and at B1 merge to a constant value after ~20 ps. This result shows that the electrons leave $1P_e$ in 20 ps but that most, ~80%, do not relax to $1S_e$ because B1 does not recover. The response may be interpreted as being inhomogeneous with different dynamics for different dots, where 80% of $1P_e$ electrons undergo rapid recombination/trapping and 20% relax to $1S_e$ within 20 ps. Alternatively, the data can be fit to a three-level system model with decay from $1P_e$ via recombination/trapping (23 ps) and intraband relaxation (74 ps) (*21*). The fast recombination/trapping of the electron in $1P_e$ is consistent with its poor stability in $1S_e$ for this ZnSe-terminated surface.

In contrast, with an outer monolayer of CdS terminated with amine ligands, the $1S_e$ electron is much more stable but IR pumping only shows a very weak and prompt $\Delta\alpha/\alpha$, ~7% maximum for B1 and within the noise at A1. This result shows that the $1P_e$ electron is short-lived—shorter than the laser pulse width of 6 ps—and it also shows that $1P_e$ completely relaxes to $1S_e$. For this system, the low-valence band of CdS confines the hole into the core + shell, and the hole is not trapped by amine ligands. This lack of trapping reduces the decoupling between electron and holes and leads to bright photoluminescence, at least for moderate shell thicknesses. The very fast relaxation confirms the importance of decoupling electrons and holes to observe slow electron cooling.

Our focus here is on the QDs that exhibit slow electron cooling. These are capped by a monolayer of CdSe treated with dodecanethiol (DDT) hole-trapping ligands. For this system, the fluorescence is quenched, as should be the case with well-decoupled electrons and holes, but the intraband spectrum and bleach α are also strong, indicating that CdSe termination, like CdS, removes electron traps. As shown in Fig. 3B, for this system, $\Delta\alpha/\alpha$ at B1 is large, typically 50%, and long-lived. A1 and B1 have similar kinetics, but the magnitude for A1 is much greater. This difference shows that the electron stays in $1P_e$ for a long time and relaxes to $1S_e$ with no intermediate state. We note that in all samples the relaxation is not a single exponential. In Fig. 3B, 60% decays with a time constant of 1.7 ± 0.3 ns, and 40% decays in 30 ± 5 ps. This range of kinetics might be caused by sample inhomogeneity because the response is so sensitive to the outer surface. We think that it could also be caused by the inhomogeneous width of the intraband absorption, across which the sample molecular IR absorption varies by one or two orders of magnitude.

We now discuss the role of ligand vibrations. Consider an intraband transition with a dipole p and an IR-absorbing surface. A dipole-coupling model gives an energy transfer time constant T_{nr},

$$\frac{1}{T_{nr}} = \frac{\lambda p^2 (2\varepsilon_0 + \varepsilon_1)^2}{24\pi^2 (2\varepsilon_0^2 + \varepsilon_1^2)\hbar R^6}\,\sigma \qquad (1)$$

(*14*), where R is the outer radius of the QD, λ is the wavelength, σ is the infrared cross section at the intraband transition caused by the surface ligands, ε_0 is the medium dielectric constant, ε_1 is the material dielectric constant, and $\hbar$ is Planck's constant divided by 2π. This model predicts a linear relation between relaxation rate and IR absorption σ. To date, alkane thiols and alkyl amines are the ligands with the lowest IR absorbance that we have used, and they always show the slowest relaxation times. Phosphonic or carboxylic acid ligands always lead to increased IR absorption (from broad hydrogen-bonding vibrations) and faster relaxation. Even mild exposure of a slow sample to these acid ligands leads to faster relaxation (Fig. 4A). Starting with a thiol-capped sample where A1 and B1 follow biexponential relaxation with a slow relaxation time of ~180 ps, a room-temperature exposure to stearic acid increases by a factor of ~7 the IR absorbance of the sample in the vicinity of the intraband transition. Correspondingly, the electron now relaxes from $1P_e$ to $1S_e$ in ~12 ps. The low magnitude of the long tail seen in the A1 and B1 response also shows that it mostly returns to $1S_e$, indicating small trapping. This correlation between increased ligand absorbance and faster relaxation supports the role of vibrations in the cooling dynamics.

Equation 1 further predicts that if the electron-confining shell is grown with similar ligand coverage, σ will increase as R^2, such that T_{nr} will

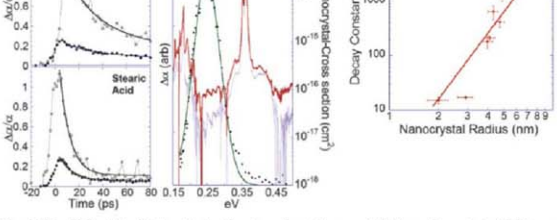


Fig. 4. Sensitivity of the intraband relaxation dynamics to the sample IR absorption and shell thickness. **(A)** CdSe/ZnS/$ZnSe_{BML}$/CdSe/DDT. A1 and B1 show the $1P_e$ decay with a slow time constant of 180 ± 40 ps. After exposure to stearic acid, A1 and B1 show faster recovery. The intraband spectrum of the sample (linear scale) is shown with solid dots and Gaussian line fit. The infrared absorption of the sample (log scale) is shown before stearic acid exposure (dotted line) and after exposure (solid line). **(B)** Slow relaxation time constant plotted against the outer radius for CdSe/ZnS/ZnSe/CdSe/thiolate of different ZnSe shell thicknesses. The line is an R^4 fit.

increase as R^4. Figure 4B shows data consistent with this predicted scaling. The slow time constant increases from ~10 ps to ~1.7 ns for samples with similar intraband energy (kept within 27 ± 3 meV for all the samples) but with shell thickness increasing from 0 to ~3.5 nm. Finally, the absolute magnitudes of lifetimes also agree with predictions from the equation. Using $p = 0.3eR_0$ (where R_0 is the core radius) and $\sigma \sim 2.10^{-18}$ cm^2, measured in the 0.28-eV region for thiol-passivated CdSe QDs with $R = R_0 = 2.25$ nm, yields $T_{IR} \sim 23$ ps. A shell 3.25 nm thick, for a total radius of 5.5 nm, would give $T_{IR} \sim 1$ ps, in agreement with the data in Fig. 3B. Thus, it is likely that the vibrational coupling mechanism is now dominating and that even longer lifetimes will be achieved by using ligands with lower IR activity, more complete exchange procedures, or fully inorganic matrices.

Slow electron cooling had long been expected in quantum dots, but it had remained elusive in colloidal dots as well as in those grown by molecular beam epitaxy (*22*–*24*). Colloidal quantum dots allow a strong confinement, which should provide a good decoupling between the electronic states and the lattice vibrations. However, to observe slow (>1 ns) electron cooling, we had to use core/shell quantum dots designed to reduce several other ways by which energy is dissipated: via coupling to holes, intermediate trap states, and molecular vibrations. The cooling time could be tuned from less than 6 ps to longer than 1 ns with identical cores and different shell compositions and thicknesses. For emitting or detecting IR radiation, the increase in cooling time by three orders of magnitude opens the prospect of using the intraband transitions of colloidal dots. Proposals to attempt to extract the energy of hot carriers from quantum dots to improve photovoltaic yield are also better validated.

References and Notes

1. X. Michalet *et al.*, *Science* **307**, 538 (2005).
2. V. I. Klimov, D. W. McBranch, *Phys. Rev. Lett.* **80**, 4028 (1998).
3. V. I. Klimov, D. W. McBranch, C. A. Leatherdale, M. G. Bawendi, *Phys. Rev. B* **60**, 13740 (1999).
4. A. J. Nozik, *Physica E* **14**, 115 (2002).
5. H. Benisty, C. M. Sotomayortorres, C. Weisbuch, *Phys. Rev. B* **44**, 10945 (1991).
6. U. Bockelmann, G. Bastard, *Phys. Rev. B* **42**, 8947 (1990).
7. T. Inoshita, H. Sakaki, *Phys. Rev. B* **46**, 7260 (1992).
8. T. Inoshita, H. Sakaki, *Physica B* **227**, 373 (1996).
9. A. L. Efros, V. A. Kharchenko, M. Rosen, *Solid State Commun.* **93**, 281 (1995).
10. L. W. Wang, M. Califano, A. Zunger, A. Franceschetti, *Phys. Rev. Lett.* **91**, 056404 (2003).
11. E. Hendry *et al.*, *Phys. Rev. Lett.* **96**, 057408 (2006).
12. P. Guyot-Sionnest, M. Shim, C. Matranga, M. Hines, *Phys. Rev. B* **60**, R2181 (1999).
13. V. I. Klimov, A. A. Mikhailovsky, D. W. McBranch, C. A. Leatherdale, M. G. Bawendi, *Phys. Rev. B* **61**, R13349 (2000).
14. P. Guyot-Sionnest, B. Wehrenberg, D. Yu, *J. Chem. Phys.* **123**, 074709 (2005).
15. D. F. Schroeter, D. J. Griffiths, P. C. Sercel, *Phys. Rev. B* **54**, 1486 (1996).
16. P. C. Sercel, *Phys. Rev. B* **51**, 14532 (1995).
17. Y. Toda, O. Moriwaki, M. Nishioka, Y. Arakawa, *Phys. Rev. Lett.* **82**, 4114 (1999).
18. Y. A. Yang, H. M. Wu, K. R. Williams, Y. C. Cao, *Angew. Chem. Int. Ed.* **44**, 6712 (2005).
19. A. Pandey, P. Guyot-Sionnest, *J. Chem. Phys.* **127**, 104710 (2007).
20. P. Reiss, J. Bleuse, A. Pron, *Nano Lett.* **2**, 781 (2002).
21. See supporting material on *Science* Online.
22. S. Sauvage *et al.*, *Phys. Rev. Lett.* **88**, 177402 (2002).
23. O. Verzelen, G. Bastard, R. Ferreira, *Phys. Rev. B* **66**, 081308 (2002).
24. J. Urayama, T. B. Norris, J. Singh, P. Bhattacharya, *Phys. Rev. Lett.* **86**, 4930 (2001).
25. Supported by NSF grant DMR-0706268 and by NSF Materials Research Science and Engineering Centers program grant DMR-0213745.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/929/DC1
Materials and Methods
Figs. S1 to S6

30 April 2008; accepted 6 October 2008
10.1126/science.1159832

Reaction-Driven Restructuring of Rh-Pd and Pt-Pd Core-Shell Nanoparticles

Feng Tao,[1,2] Michael E. Grass,[1,2] Yawen Zhang,[1,2,5] Derek R. Butcher,[1,2] James R. Renzas,[1,2] Zhi Liu,[1,3] Jen Y. Chung,[3] Bongjin S. Mun,[3] Miquel Salmeron,[1,4]* Gabor A. Somorjai[1,2]*

Heterogeneous catalysts that contain bimetallic nanoparticles may undergo segregation of the metals, driven by oxidizing and reducing environments. The structure and composition of core-shell $Rh_{0.5}Pd_{0.5}$ and $Pt_{0.5}Pd_{0.5}$ nanoparticle catalysts were studied in situ, during oxidizing, reducing, and catalytic reactions involving NO, O_2, CO, and H_2 by x-ray photoelectron spectroscopy at near-ambient pressure. The $Rh_{0.5}Pd_{0.5}$ nanoparticles underwent dramatic and reversible changes in composition and chemical state in response to oxidizing or reducing conditions. In contrast, no substantial segregation of Pd or Pt atoms was found in $Pt_{0.5}Pd_{0.5}$ nanoparticles. The different behaviors in restructuring and chemical response of $Rh_{0.5}Pd_{0.5}$ and $Pt_{0.5}Pd_{0.5}$ nanoparticle catalysts under the same reaction conditions illustrates the flexibility and tunability of the structure of bimetallic nanoparticle catalysts during catalytic reactions.

The development of bimetallic nanoparticles (NPs) with controlled size, composition, and structure opens enormous possibilities for engineering catalysts with enhanced activity and selectivity. Important technological areas, including catalytic reforming (*1*, *2*), pollution control (*1*), alcohol oxidation (*3*), and electrocatalysis in fuel cells (*4*), are based on bimetallic catalytic systems (*5*, *6*). Although it is known that the structure and composition of the surface of materials can be modified in response to changing reaction conditions (*1*, *7*, *8*), when the size of the material is at the nanometer scale, the changes can be much more dramatic as the distinction between surface and bulk regions fades away. In the case of NP catalysts, these changes can transform the material so that it has distinctive reactivity.

We demonstrated that bimetallic NP catalysts can undergo profound structural and chemical changes in response to reactive environments at ambient pressures. This was made possible by the use of an ambient-pressure x-ray photoelectron spectroscopy (APXPS) apparatus (*9*, *10*) that can be used to obtain XPS data from materials exposed to gas pressures much higher than the usual limits imposed by high-vacuum conditions, recently up to ~5 to 10 torr. With it, we could study in situ the structure and composition of core-shell $Rh_{0.5}Pd_{0.5}$ and $Pt_{0.5}Pd_{0.5}$ bimetallic NPs during catalytic reactions in different gas environments. The composition and distribution of the constituent elements within the shell of $Rh_{0.5}Pd_{0.5}$ NPs were found to change dramatically in response to changes in reactant gas composition, a result that demonstrates the structural flexibility of NPs and the interplay of structure and reactivity.

$Rh_{0.5}Pd_{0.5}$ and $Pt_{0.5}Pd_{0.5}$ bimetallic NPs with diameters of 15 ± 2 nm were synthesized by means of colloidal chemistry methods [see supporting online material (SOM) (*11*)] and characterized by transmission electron microscopy and x-ray diffraction. NPs ($Rh_{0.5}Pd_{0.5}$ or $Pt_{0.5}Pd_{0.5}$) were deposited on the oxidized surface of silicon wafers to form model catalysts for XPS studies. The spectra were obtained at x-ray photon energies of 645 and 850 eV at the Advanced Light Source at the Lawrence Berkeley National Laboratory (beamline 9.3.2), the Berkeley Synchrotron Facility, and in ultrahigh vacuum (UHV) using Al Kα x-rays at 1486.6 eV in a laboratory Phi 550 instrument. The mean free paths (MFPs) of Rh3d and Pd3d photoelectrons excited at these

[1]Materials Sciences and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [3]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA. [5]College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

*To whom correspondence should be addressed. E-mail: somorjai@berkeley.edu (G.A.S.); mbsalmeron@lbl.gov (M.S.)

three x-ray energies were approximately 0.7, 1.0, and 1.6 nm, respectively (*12*). The structure of the $Pt_{0.5}Pd_{0.5}$ NPs was studied at photon energies of 630, 880, and 1486.6 eV for Pd3d, and at 350, 630, and 1486.6 eV for Pt4f. The MFPs of the generated Pd3d and Pt4f photoelectrons were also 0.7, 1.0, and 1.6 nm, respectively (*12*).

Before their use in catalytic reactions, the as-synthesized $Rh_{0.5}Pd_{0.5}$ NPs were found to be Rh-rich in the surface layers (Fig. 1A). At the lowest photon energy [$h\nu$ = 645 eV, kinetic energy (KE) = ~335 eV], corresponding to a MFP of ~0.7 nm, the measured atomic fraction of Rh was 0.93 ± 0.03. The determination of atomic fractions of Rh and Pd is described in the SOM (*11*). The average Rh atomic fraction decreased to 0.86 ± 0.03 and 0.52 ± 0.03 at depths corresponding to MFPs of 1.0 nm ($h\nu$ = 850 eV, KE = ~540 eV) and 1.6 nm ($h\nu$ = 1486.6 eV, KE = ~1180 eV), respectively. Here we use the term "shell" to refer to the three to four atomic layers within 0.7 nm of the surface. The section between 0.7 and 1.6 nm is referred to as the "intermediate layer," and the section from 1.6 nm to the center of a NP is referred to as the "central core." The volume ratio of material from the surface to the boundary of each of these three regions is respectively 25, 35, and 51%.

$Pt_{0.5}Pd_{0.5}$ NPs have a similar core-shell structure, but the shell is Pd-rich (Fig. 1B). The observed atomic fraction of Pd in these NPs was 0.84 ± 0.03, 0.67 ± 0.03, and 0.52 ± 0.03 within the MFP distances of 0.7, 1.0, and 1.6 nm, respectively.

After the initial characterization, we measured the changes in surface composition and chemical state of the $Rh_{0.5}Pd_{0.5}$ NPs under oxidizing (100 mtorr NO or O_2), catalytic (100 mtorr NO and 100 mtorr CO reacting to produce N_2 and CO_2), and reducing (100 mtorr CO or H_2) conditions using APXPS. The atomic fractions presented in Fig. 2 were obtained with an x-ray energy of 645 eV and thus represent the changes in composition in the 0.7-nm shell. The top part of Fig. 2 shows a substantial oscillation of the relative atomic fractions as the gas environment changed from oxidizing to catalytic at 300°C. After oxidation by 100-mtorr NO, the Rh in the shell was almost completely oxidized, with ~94% of the Rh in oxide form. When 100-mtorr CO was added to the 100-mtorr NO to produce CO_2 and N_2 (catalytic conditions), the total Rh atomic fraction in the shell decreased from 0.92 ± 0.03 to 0.46 ± 0.02 and that of Pd increased from 0.08 ± 0.03 to 0.54 ± 0.02 (reaction 2 in top part of Fig. 2). This result indicates that a drastic restructuring of the shell and core of the NP took place, where Pd migrated to the shell and Rh migrated to the core. The production of CO_2 and N_2 formed in the reaction was measured by sampling gas into a mass spectrometer located in a separate UHV chamber, indicating that the NPs were active during CO and NO catalytic conversion. The ~1.0-eV downshift in binding energy of the Rh3d core level indicates the reduction of RhO_y to metal Rh^0. The Rh that remained in the shell was ~76% metallic during the catalytic reaction (reaction 2), in contrast to the ~94% RhO_y during the oxidizing reaction. The reduction of RhO_y to metallic Rh shows a substantial chemical response accompanying the change in atomic distribution.

The observed changes in atomic distribution and chemical state are reversible and depend on the composition of the surrounding reactive gases, as shown in a sequence of five reactions in Fig. 2. If CO is removed while NO remains in the reactor chamber (reaction 3 in Fig. 2), Rh diffuses back to the shell and becomes substantially oxidized. Analysis of the XPS peaks shows that the reconstructed shell contains ~72 ± 3% Rh (top part of Fig. 2), of which ~90% is oxidized (bottom panel). If CO is introduced again, added to the NO gas phase (reaction 4), the spatial distribution and chemical state of the atoms are restored to those during reaction 2. The restructuring of the shell and core of the $Rh_{0.5}Pd_{0.5}$ NPs can be repeated several times by changing the gas composition from oxidizing (NO or O_2) to reducing (CO) or catalytic (CO + NO) conditions as shown in Fig. 2 and fig. S3 (*11*). If another reducing gas, hydrogen, is used instead [reaction 2 in fig. S4 (*11*)], the RhO_y is reduced, with the Rh atoms migrating back to the core and the Pd atoms to the shell, forming a Pd-rich shell

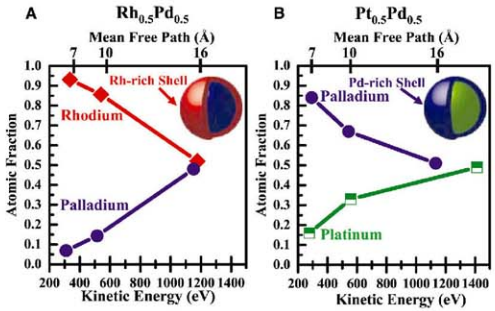


Fig. 1. (A) Dependence of Rh and Pd atomic fractions of as-synthesized $Rh_{0.5}Pd_{0.5}$ NPs measured at 25°C in UHV as a function of photoelectron KE and mean free path. Detected photoelectrons with a KE of ~1152 eV originate from the shell, intermediate layer, and central core of the $Rh_{0.5}Pd_{0.5}$ NPs. The signal of photoelectrons with KE of ~335 and ~540 eV is mainly contributed by the atoms in surface layers because of the shorter MFP of these photoelectrons. **(B)** Dependence of Pd and Pt atomic fractions of the as-synthesized $Pt_{0.5}Pd_{0.5}$ measured in UHV at 25°C as a function of the KE and mean free path of the excited photoelectrons. Schematics showing the core-shell structures of the $Rh_{0.5}Pd_{0.5}$ and $Pt_{0.5}Pd_{0.5}$ NPs are included (these schematics do not represent the shape of the NPs). The *y*-axis data points have an associated error of ±0.03.

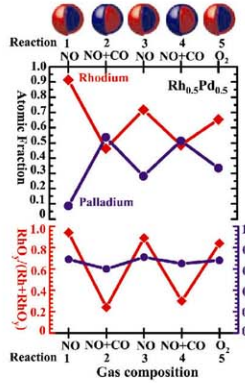


Fig. 2. (Top) Evolution of Rh (Rh^0 + Rh^{2y+}) and Pd (Pd^0 + Pd^{2y+}) atomic fractions in the $Rh_{0.5}Pd_{0.5}$ NPs at 300°C under oxidizing conditions (100 mtorr NO or O_2) and catalytic conditions (100 mtorr NO and 100 mtorr CO) denoted in the *x* axis. **(Bottom)** Evolution of the fraction of the oxidized Rh (left *y* axis) and Pd atoms (right *y* axis) in the examined region under the same reaction conditions as the top part of the figure. All atomic fractions in this figure were obtained with an x-ray energy of 645 eV for Rh3d and Pd3d, which generates photoelectrons with a MFP of ~0.7 nm. Schematic diagrams above the top of the figure show the reversible segregation of Rh and Pd under alternating oxidizing and catalytic conditions. The *y*-axis data points for reactions 1, 3, and 5 have an associated error of ±0.03; for reactions 2 and 4, the error bar is ±0.02.

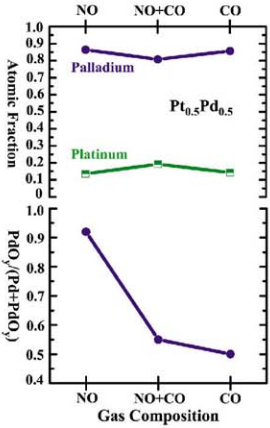


Fig. 3. (**Top**) Evolution of the Pd and Pt atomic fractions in $Pt_{0.5}Pd_{0.5}$ NPs at 300°C under oxidizing (100 mtorr NO), catalytic (100 mtorr NO and 100 mtorr CO), and reducing (100 mtorr CO) conditions. The *x* axis represents the different gas environments. (**Bottom**) Evolution of the atomic fraction of the oxidized Pd atoms in the examined region under the same reaction conditions as the top panel. All atomic fractions were obtained at an x-ray energy of 350 eV for Pt4f (KE ~280 eV) and 630 eV for Pd3d (KE ~ 280 eV). The *y*-axis error in the data is ±0.03.

similar to that formed in pure CO [fig. S3 (*11*)]. When Al Kα x-rays at 1486.6 eV were used, the Rh fraction in the $Rh_{0.5}Pd_{0.5}$ NPs measured in UHV after alternating oxidizing and reducing reactions [fig. S4 (*11*)] was 0.36 ± 0.03, much lower than the 0.52 ± 0.03 value of the as-synthesized NPs before reaction also measured in UHV with Al Kα x-rays (Fig. 1A). This difference indicates that the core region (at depth greater than 1.6 nm) also participates in the restructuring of the NPs.

The opposite segregation behavior of Rh and Pd under oxidizing and reducing conditions can be explained by considering the surface energy in the metals and in the oxides. The lower surface energy of Pd relative to Rh tends to drive Pd metal atoms to the surface (*13*–*15*). The fact that the Rh oxide is more stable than the Pd oxide provides the driving force for the segregation and preferential oxidation of Rh at the surface (*16*). When a reducing gas, CO, is added to NO, the oxides are reduced to the metallic state and the oxygen atoms react with adsorbed CO to form CO_2 and desorb. Because of their higher surface free energy, the Rh atoms migrate to the core, thereby decreasing the atomic fraction of Rh in the shell under reducing and catalytic conditions.

In the $Pt_{0.5}Pd_{0.5}$ NPs, synthesized with the same procedure described in the SOM, the shell region is substantially richer in Pd. Pd has lower surface energy and is a more reactive metal than Pt (*15*, *16*). Under oxidizing conditions, it forms a shell with 92% PdO_y without substantial segregation of Pt atoms as compared to the as-synthesized $Pt_{0.5}Pd_{0.5}$ NPs (Figs. 3 and 1B). During catalytic reaction and under reducing conditions, the PdO_y is substantially reduced (bottom part of Fig. 3). There is no obvious segregation under these conditions. Compared to $Rh_{0.5}Pd_{0.5}$, however, the $Pt_{0.5}Pd_{0.5}$ NPs do not exhibit the strong segregation and reversibility characteristic of the $Rh_{0.5}Pd_{0.5}$ NPs as the gas composition changes sequentially from oxidizing, to catalytic, to reducing as shown in the top part of Fig. 3. Because Pt is much less easily oxidized (*16*) and has higher surface energy (*15*), the Pt atoms are not pinned to the surface by the formation of oxide. Thus, there is no substantial atomic reorganization in reactive environments.

The restructuring phenomenon observed in the bimetallic NPs induced by changes in reactive gas offers an interesting way of engineering the nanostructure of NPs for catalysis and other applications. One goal could be the synthesis of "smart" catalysts whose structure changes advantageously depending on the reaction environment. Our results suggest that the combination of a tunable colloid chemistry-based synthesis, followed by the controllable engineering of the structure of NPs with the use of reactive gases, opens a new door for designing new catalysts and shaping the catalytic properties of nanomaterials by structural engineering in reactive environments.

References and Notes

1. G. A. Somorjai, *Introduction to Surface Chemistry and Catalysis* (Wiley, New York, 1994), pp. 500–512.
2. G. W. Huber, J. W. Shabaker, J. A. Dumesic, *Science* **300**, 2075 (2003).
3. D. I. Enache *et al.*, *Science* **311**, 362 (2006).
4. F. Maroun, F. Ozanam, O. M. Magnussen, R. J. Behm, *Science* **293**, 1811 (2001).
5. C. Burda, X. Chen, R. Narayanan, M. A. El-Sayed, *Chem. Rev.* **105**, 1025 (2005).
6. A. Christensen, P. Stoltze, J. K. Norskov, *J. Phys. Condens. Matter* **7**, 1047 (1995).
7. C. T. Campbell, *Annu. Rev. Phys. Chem.* **41**, 775 (1990).
8. J. A. Rodriguez, D. W. Goodman, *Science* **257**, 897 (1992).
9. D. F. Ogletree *et al.*, *Rev. Sci. Instrum.* **73**, 3872 (2002).
10. M. Salmeron, R. Schlögl, *Surf. Sci. Rep.* **63**, 169 (2008).
11. Methods and additional data are available as supporting material on *Science* Online.
12. C. J. Powell, A. Jablonski, *NIST Electron Inelastic-Mean-Free-Path Database*, ed. 1.1 (National Institute of Standards and Technology, Gaithersburg, MD, 2000).
13. A. V. Ruban, H. L. Skriver, J. K. Norskov, *Phys. Rev. B* **59**, 15990 (1999).
14. T. Maillet, J. Barbier, P. Gelin, H. Praliaud, D. Duprez, *J. Catal.* **202**, 367 (2001).
15. H. L. Skriver, N. M. Rosengaard, *Phys. Rev. B* **46**, 7157 (1992).
16. M. Salmeron, L. Brewer, G. A. Somorjai, *Surf. Sci.* **112**, 207 (1981).

17. Supported by the director of the Office of Science; Office of Advanced Scientific Computing Research; Office of Basic Energy Sciences, Materials Sciences and Engineering; and Chemical Sciences, Geosciences, and Biosciences Division of the U.S. Department of Energy under contract DE-AC02-05CH11231.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164170/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
References

4 August 2008; accepted 29 September 2008
Published online 9 October 2008;
10.1126/science.1164170
Include this information when citing this paper.

Reconstructing Farallon Plate Subduction Beneath North America Back to the Late Cretaceous

Lijun Liu,* Sonja Spasojević, Michael Gurnis

Using an inverse mantle convection model that assimilates seismic structure and plate motions, we reconstruct Farallon plate subduction back to 100 million years ago. Models consistent with stratigraphy constrain the depth dependence of mantle viscosity and buoyancy, requiring that the Farallon slab was flat lying in the Late Cretaceous, consistent with geological reconstructions. The simulation predicts that an extensive zone of shallow-dipping subduction extended beyond the flat-lying slab farther east and north by up to 1000 kilometers. The limited region of flat subduction is consistent with the notion that subduction of an oceanic plateau caused the slab to flatten. The results imply that seismic images of the current mantle provide more constraints on past tectonic events than previously recognized.

Mantle convection is a time-dependent process, as evident through the evolution of plate motions, changes in the configuration of plate margins, and pulses of hotspot volcanism. How does convection actually change and what time-dependent models are consistent with the range of observational constraints? Answers to these questions are important because the time dependence of the system is strongly dependent on key uncertainties, including the depth dependence of mantle viscosity, the rheology and buoyancy of slabs, and the thermo-

dynamics of solid-solid phase transitions (*1–3*). Modeling past convection accurately enough to reveal causes of geological and current seismic data is difficult because forward models are strongly dependent on uncertain initial conditions such as the mantle's thermal structure. This ambiguity of initial conditions, mantle viscosity, and the link between time-dependent models and observations can be partially overcome with inverse models of mantle convection, as we demonstrate with a model tailored to the evolution of North America since the Late Cretaceous.

Our models use seismic tomography that resolves the remnants of Farallon plate subduction since the Mesozoic (*4–6*) as input for defining present-day structure (*7*). We invert mantle convection with the adjoint method that constrains the initial condition iteratively (*8–10*). Plate motions (*11*) are imposed as kinematic boundary conditions allowing the mapping between dynamic topography and stratigraphic observations (*12*). Because the temporal evolution of surface dynamic topography is related to poorly known mantle buoyancy and viscosity (*9*), we constrain these uncertainties with stratigraphic data, including reconstructions of paleoshorelines (*13*), sediment isopachs (*14*), and borehole tectonic subsidence (*15*, *16*).

A standard model of mantle convection (*17*) generates a first guess of the initial condition with a simple backward integration from the present to the past by reversing the sign of gravity, during which the Farallon slab remnants rise (instead of sink) (Fig. 1A). By 70 million years ago (Ma), within the upper mantle, the cold mantle diverges and partly deflects toward the east as a result of the reversely imposed plate motions (Fig. 1A), yielding a geophysically unreasonable subduction configuration. This cannot be overcome either by varying the radial viscosity structure or by performing additional forward-adjoint iterations, and such models are inconsistent with the temporal sequence of subsidence and uplift of the western interior (*18*, *19*). Essentially, the present Farallon seismic anomaly is too far to the east to be simply connected to the Farallon-North American boundary in the Mesozoic, a result implicit in forward models (*20*).

When we incorporate a simple parameterized stress guide (*21*) beneath the North American plate, the cold material becomes preferentially attached to the Farallon plate. Through adjoint iterations, the position and structure of the slab further adjust to reasonable subduction geometries (Fig. 1B). The stress guide strongly influences slab motion horizontally in the upper mantle, whereas lower mantle flow is less affected (Fig. 1, A and B). As the cold slab material rises in a backward sense, the upper part of the slab moves faster horizontally toward the trench than the lower part, and this leads to the formation of a flat- to shallow-dipping slab around 70 Ma extending 2000 km inland from the trench (Fig. 1B).

For a given density, the timing and extent of the flat slab formation depend on mantle viscosities, especially lower mantle viscosity, because the slab has a longer path in the lower compared to the upper mantle. The surface dynamic topography associated with the descending slab causes the continent to subside dynamically (*18*, *19*) and results in marine deposition, if below sea level. In practice, we assume that the sedimentation rate was constant and sea level eustatic while isostatically adjusting basin depth after each time step of 1 million years. The observed extent of Cretaceous flooding (Fig. 2A) (*14*, *22*), therefore, provides direct constraint on mantle viscosity. In a backward sense, a high viscosity (3×10^{22} Pa s) lower mantle causes the slab to rise slowly, reach the surface closer to the trench, and form a small area of flat slab subduction; the associated dynamic subsidence causes flooding with a smaller wavelength over shorter duration than observed (Fig. 2B). Alternatively, a mantle of smaller viscosity causes the slab to rise faster and reach the surface farther from the trench while forming a larger area of flat slab subduction, which causes extensive flooding spatially and temporally (Fig. 2C). A good fit to observed flooding is obtained if we decrease the lower mantle viscosity from the case in Fig. 2B by a factor of two to 1.5×10^{22} Pa s (Fig. 2D). Tectonic subsidence extracted from boreholes reflects the magnitude and duration of surface subsidence for points fixed relative to the continent, and it is controlled by both the effective temperature and mantle rheology. Observed flooding and borehole subsidence jointly constrain the amplitude of the temperature anomaly, which, if too large, overpredicts either flooding or tectonic subsidence or

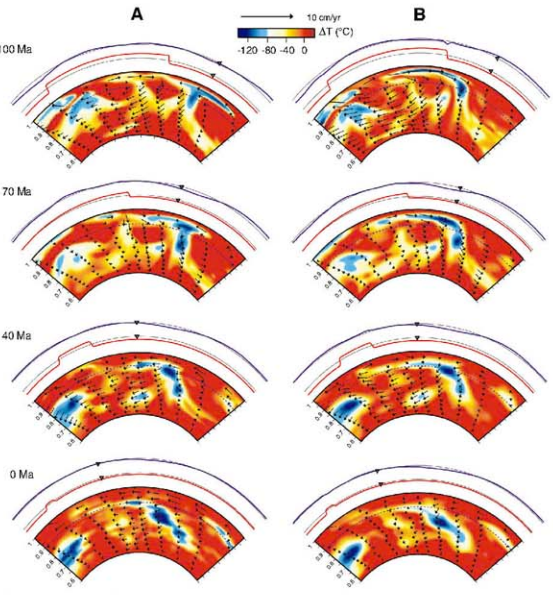


Fig. 1. (A) Temporal evolution of the Farallon slab during the simple backward integration at latitude 41°N with a standard model (*17*), where the lower:upper mantle viscosity is 15:1 (relative to 10^{21} Pa s) and the maximum effective temperature anomaly is 160°C. **(B)** Same as (A) but for a forward model based on the initial condition recovered with the adjoint method incorporating a stress guide (*21*). Cross sections show temperature anomalies, velocity vectors (in the forward sense), dynamic topography (blue) and longitudinal component of plate motion (red, positive is eastward) along this profile. The black triangles indicate the position of a borehole site (41.6°N, 254°E) about 250 km north of Denver. The dashed blue line marks upper-lower mantle interface. At 0 Ma, the left panel is seismic tomography, but the right is the prediction of tomography based on the forward-adjoint looping.

Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: lijun@gps.caltech.edu

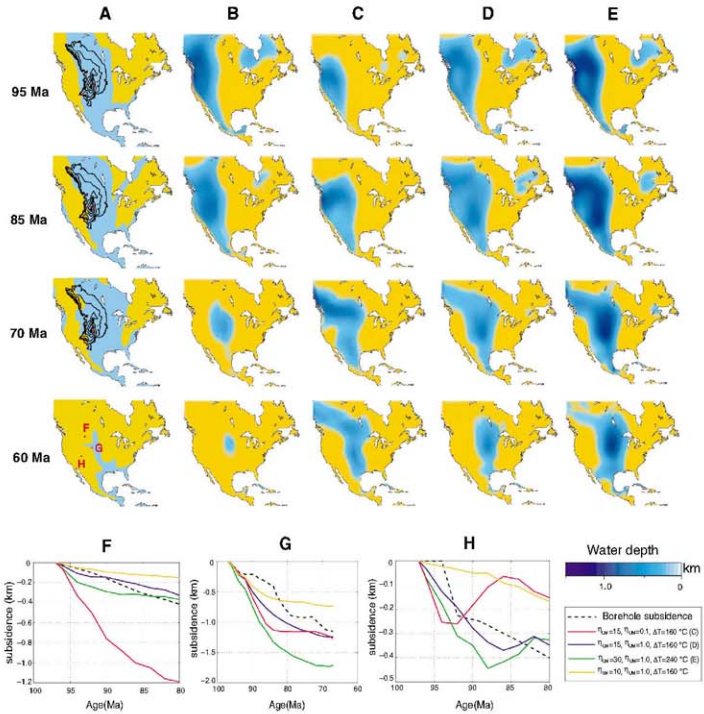


Fig. 2. Observed and predicted continental flooding and borehole subsidence for different temperature scaling and mantle viscosities. **(A)** Geologically inferred flooding (blue areas) bounded by paleoshorelines (*13*) with cumulative Cretaceous isopachs (*14*) overlain (2-km contour interval), with red dots indicating the three boreholes shown in (F) to (H). **(B to E)** Predicted flooding with lower-mantle viscosity η_{LAM} = 30, 15, 30 and upper-mantle viscosity η_{UAM} = 1, 0.1, 1, 1, respectively (relative to 10^{21} Pa s); cases (B) to (D) all have an effective temperature magnitude ΔT = 160°C, and (E) has ΔT = 240°C, where (D) is the best-fit model. **(F to H)** Observed and predicted borehole dynamic subsidence from four different models.

both (Fig. 2, E to H). An unreasonably small temperature anomaly does not reproduce as widespread flooding as observed (Fig. 2A). Borehole subsidence rate is inversely proportional to the upper mantle viscosity (*9*), and models constrained with such data show that decreasing upper mantle viscosity increases the rate of tectonic subsidence (Fig. 2, G and H). The phase of the borehole subsidence is another constraint: With data from three boreholes about 1000 km (north-south) and 800 km (east-west) apart, we can constrain the shape and position of the Farallon slab.

Our preferred model (Fig. 2D), which fits both flooding and borehole subsidence simultaneously, has a lower mantle viscosity of 1.5 × 10^{22} Pa s, an upper mantle viscosity of 1 × 10^{21} Pa s, and a present-day effective temperature (*21*) anomaly of 160°C. The correlation of water depth and dynamic topography with sediment isopachs (Fig. 2A) is used to further qualitatively validate the inversions. The inferred viscosities fall within bounds obtained with postglacial rebound (*23*), providing confirmation of the method. Given the simplicity of our model that has a three-layer viscosity mantle, limited spatial resolution of seismic tomography, and uncertainties in plate tectonic reconstructions, the inferred mantle properties are subject to further refinement.

Our preferred model depicts the evolution of the Farallon subduction from its early stage and includes an episode of flattening (Fig. 3). At 100 Ma, the slab is shallow-dipping, but not flat-lying (Fig. 3A). Slab flattening starts at about 90 Ma progressively from the trench and is characterized by a patch of thickened oceanic lithosphere about 1000 km wide that subducts to the northeast (Fig. 3B). The model predicts that this segment of the Farallon slab evolved into an isolated and thickened patch of oceanic lithosphere bounded by a series of shallow- to steep-dipping fragments (Fig. 3D). This configuration is sustained throughout the Late Cretaceous, accompanying a vast inland migration of the whole slab. In the model, the flat-lying slab sinks vertically from 70 Ma (Fig. 1B) (*21*). The inferred position of flat-lying subduction from the adjoint model correlates well with the region of basement-cutting Laramide-type faults in the western United States from 80 to 65 Ma (*24*). Our best model matches the along-strike extent of the faulting and its maximum in board extent (Fig. 3, B to D). However, both the onset and demise of flat subduction occur 10 million years earlier in the model compared with geological data. We made no attempt to tune model parameters based on the geologically inferred position of flat subduction. A present-day analog is the South American flat subduction in Peru, with its proposed linkage to subduction of an oceanic plateau (*25*). The morphology of the flat slab in our model (Fig. 3) may be consistent with a plateau subduction (*26*), showing a patch that is thicker and shallower than surrounding slab segments (Fig. 3).

Our inversion reverses subduction to only ~100 Ma, but even at that time, a shallow-dipping segment is intruding beneath the conti-

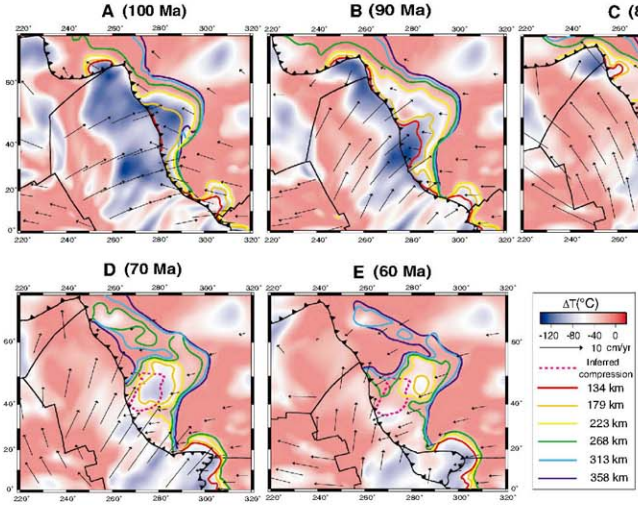


Fig. 3. Map view evolution of the Farallon subduction from the Late Cretaceous (**A** to **D**) to the Early Tertiary (**E**). The background temperature is at a depth of 223 km, and velocity vectors indicate imposed plate motions. The color lines (isotherms for a temperature of 30°C lower than ambient mantle) indicate the upper boundary of the slab at different depths with a warm (cold) color representing a small (large) depth. The dashed pink contour outlines the geologically inferred surface compression area occurring between 80 and 65 Ma (*24*).

nent (Fig. 3A). Consequently, because the inversion goes back to the period of anomalous subduction but no earlier, the conditions that gave rise to the flat-lying subduction cannot be recovered. Essentially, it is the present-day mantle structure that returns the flat-lying slab, as is evident from the predicted three-dimensional evolution of the Farallon plate subduction (*21*).

The adjoint models suggest that a vast zone of shallow-dipping subduction extended more than 1000 km eastward and northward from the zone of flat-lying subduction, especially from 90 to 70 Ma (Fig. 3, B to D). The zone is larger than inferred from a simplified set of forward models that neither incorporated the details of subduction nor attempted to match stratigraphy (*20*). The region of dynamic subsidence for North America, previously thought to be confined to portions of the western United States (*18*, *19*), was likely substantially larger in extent and mostly characterized by a Cenozoic unconformity (*19*). This relation is consistent with earlier suggestions that changes in dynamic topography should be reflected by unconformities (*19*). Much of the Canadian shield is sediment free, and the region has not been used to constrain dynamic models, but recent low-temperature thermochronology for basement samples in Northern Saskatchewan are consistent with 1 km of sediment burial and unroofing during the Cretaceous (*27*). This suggests that there is substantial ability to infer tectonic events when an adjoint of the convection problem is applied to seismic and plate motion data.

References and Notes

1. G. Schubert, D. L. Turcotte, P. Olson, in *Mantle Convection in the Earth and Planets* (Cambridge Univ. Press, Cambridge, 2001), pp. 417–498.
2. G. F. Davies, in *Dynamic Earth: Plates Plumes and Mantle Convection* (Cambridge Univ. Press, Cambridge, 1999), pp. 262–323.
3. S. Zhong, M. Gurnis, *Science* **267**, 838 (1995).
4. S. P. Grand, *Philos. Trans. R. Soc. London A* **360**, 2475 (2002).
5. R. D. Van der Hilst, S. Widiyantoro, E. R. Engdahl, *Nature* **386**, 578 (1997).
6. Y. Ren, E. Stutzmann, R. D. Van der Hilst, J. Besse, *J. Geophys. Res.* **112**, B01302 (2007).
7. Both shear (*4*) and compressional (*5*, *6*) wave tomography show a prominent positive slablike seismic anomaly at mid-mantle depths below the east coast of North America. The high-velocity component of the shear wave seismic tomography (*4*) is converted to effective temperature (*21*) with a scaling factor of 2×10^{-5} °C km^{-1} s. The upper 250-km signal associated with the North American craton is removed because it is likely neutrally buoyant (*28*). We also remove structures below 2400-km depth, where there is a clear gap in the tomographic image. For the rest, we assume a constant seismic-to-temperature scaling that we constrain by fitting our models to stratigraphic data.
8. H.-P. Bunge, C. R. Hagelberg, B. J. Travis, *Geophys. J. Int.* **152**, 280 (2003).
9. L. Liu, M. Gurnis, *J. Geophys. Res.* **113**, B08405 (2008).
10. The adjoint method, a gradient-based inversion, optimizes the initial condition by minimizing the mismatch between a model prediction and observation. For the nonlinear mantle convection, the observational constraint is seismic tomography. A simple backward integration of present-day mantle structures to the initial time provides a first-order estimate of the initial condition (*9*). In a forward-adjoint looping scheme, the forward run—which solves for conservation of mass ($u_{i,i} = 0$, where u is velocity), momentum [$-P_{,i} + (\eta u_{i,j} + \eta u_{j,i})_{,j} + \rho_0 \alpha \Delta T g \delta_{i3} = 0$, where P is dynamic pressure, η viscosity, ΔT effective temperature, α thermal expansion, ρ_0 reference density, and g gravitational acceleration], and energy ($T_{,t} + u_i T_{,i} = \kappa T_{,ii}$, where T is temperature, t time, and κ thermal diffusivity) for an incompressible Newtonian fluid—predicts the present-day mantle thermal structure, whereas the adjoint run, which solves for the kinematic adjoint equation $\Theta_{,t} + u_i \lambda_{i,j} + \kappa \lambda_{i,ii} = 0$, where λ is adjoint temperature) with u stored from the forward run—back-propagates the mismatch to the initial time. The past mantle thermal state is iteratively updated until the mismatch with the observation (tomography) is small. The adjoint method has been incorporated (*9*) in a finite-element incompressible model of mantle convection (*17*).
11. We use GPlates reconstructions of global plate motions at 1-million-year intervals, in which the plate margins continuously evolve with self-consistent velocities between the plates and plate margins (*29*). The rotation poles of (*30*) are used in the GPlates reconstructions, implemented in a moving hotspot reference frame.
12. M. Gurnis, R. D. Müller, L. Moresi, *Science* **279**, 1499 (1998).
13. A. G. Smith, D. G. Smith, B. M. Funnel, *Atlas of Mesozoic and Cenozoic Coastlines* (Cambridge Univ. Press, Cambridge, 1994).
14. T. D. Cook, A. W. Bally, *Stratigraphic Atlas of North and Central America* (Princeton, Univ. Press, Princeton, NJ, 1975).
15. M. Pang, D. Nummedal, *Geology* **23**, 173 (1995).
16. S. Liu, D. Nummedal, P.-G. Yin, H.-J. Luo, *Basin Res.* **17**, 487 (2005).
17. We use Citcom*S*, a finite-element model (*31*) on a global mesh with a 129-by-129-by-65 grid on each of the 12 caps. We assume a free slip and isothermal core-mantle boundary and an isothermal surface. A Rayleigh number

(based on thickness of the mantle) of 9.4×10^{6} is used (21). The three-layer viscosity includes a lithosphere with constant 5×10^{22} Pa s viscosity, an upper mantle, and a lower mantle with varying viscosities to be constrained. Laterally, viscosity increases exponentially by one order of magnitude for a temperature drop of 200°C.

18. J. X. Mitrovica, C. Beaumont, G. T. Jarvis, *Tectonics* **8**, 1079 (1989).
19. P. M. Burgess, M. Gurnis, L. Moresi, *Bull. Geol. Soc. Am.* **109**, 1515 (1997).
20. H.-P. Bunge, S. P. Grand, *Nature* **405**, 337 (2000).
21. Materials and methods are available as supporting material on *Science* Online.
22. G. A. Bond, *Geology* **4**, 557 (1976).
23. G. A. Milne *et al.*, *J. Geophys. Res.* **109**, B02412 (2004).
24. J. Saleeby, *Bull. Geol. Soc. Am.* **115**, 655 (2003).
25. M.-A. Gutscher, W. Spakman, H. Bijwaard, E. R. Engdahl, *Tectonics* **19**, 814 (2000).
26. R. F. Livaccari, K. Burke, A. M. C. Sengor, *Nature* **289**, 276 (1981).
27. R. M. Flowers, *Eos* **52**, abstr. V34C-06 (2007).
28. S. Goes, S. van der Lee, *J. Geophys. Res.* **107**, 2050 (2002).
29. M. Gurnis *et al.*, *Eos* **52**, abstr. DI14A-07 (2007).
30. R. D. Müller *et al.*, *Geochem. Geophys. Geosys.* **9**, 10.1029/2007GC001743 (2008).
31. S. Zhong, M. T. Zuber, L. Moresi, M. Gurnis, *J. Geophys. Res.* **105**, 11,063 (2000).
32. This is contribution no. TO 88 of the Caltech Tectonics Observatory. The work was partially supported by the Gordon and Betty Moore Foundation through the Tectonics Observatory and the NSF through EAR-0609707. We appreciate discussions with J. Saleeby and R. D. Müller. The original Citcom5 software was obtained from the Computational Infrastructure for Geodynamics (CIG) (http://geodynamics.org).

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/934/DC1
Materials and Methods
Figs. S1 to S3
References

8 July 2008; accepted 8 October 2008
10.1126/science.1162921

Lack of Exposed Ice Inside Lunar South Pole Shackleton Crater

Junichi Haruyama,[1] Makiko Ohtake,[1] Tsuneo Matsunaga,[2] Tomokatsu Morota,[1] Chikatoshi Honda,[1] Yasuhiro Yokota,[1] Carle M. Pieters,[3] Seiichi Hara,[4] Kazuyuki Hioki,[4] Kazuto Saiki,[5] Hideaki Miyamoto,[6] Akira Iwasaki,[7] Masanao Abe,[1] Yoshiko Ogawa,[2] Hiroshi Takeda,[8] Motomaro Shirao,[9] Atsushi Yamaji,[10] Jean-Luc Josset[11]

The inside of Shackleton Crater at the lunar south pole is permanently shadowed; it has been inferred to hold water-ice deposits. The Terrain Camera (TC), a 10-meter-resolution stereo camera onboard the Selenological and Engineering Explorer (SELENE) spacecraft, succeeded in imaging the inside of the crater, which was faintly lit by sunlight scattered from the upper inner wall near the rim. The estimated temperature of the crater floor, based on the crater shape model derived from the TC data, is less than ~90 kelvin, cold enough to hold water-ice. However, at the TC's spatial resolution, the derived albedo indicates that exposed relatively pure water-ice deposits are not on the crater floor. Water-ice may be disseminated and mixed with soil over a small percentage of the area or may not exist at all.

Whether or not an amount of concentrated hydrogen on the lunar poles (*1*) forms water-ice is both a scientifically intriguing issue and a potentially important research subject in order for humans to settle on the Moon and travel further into space. Possible reservoirs of hydrogen on the lunar poles are permanently shadowed areas (PSAs), which receive no direct sunlight and are extremely cold (*2*, *3*). Because the present rotation inclination of the Moon is nearly zero (~1.5° from normal to the ecliptic plane), topographic lows on the lunar poles become PSAs. Shackleton Crater, which lies at the lunar south pole, has therefore been considered as a possible water-ice reservoir in its PSA. Bistatic radar observations made by the Clementine probe (*4*, *5*) implied that there are water-ice deposits inside Shackleton Crater. However, subsequent Earth-based radar observations showed little evidence for the existence of water-ice deposits (*6*), although they could observe only an upper part of the inner wall of Shackleton.

We investigated the interior of Shackleton Crater with the panchromatic Terrain Camera (TC), a 10-m-resolution stereo camera onboard the Selenological and Engineering Explorer (SELENE) spacecraft (also nicknamed Kaguya) (*7*). The method of our observation was based on the idea that the PSA is weakly lit by sunlight scattered from nearby higher terrains. The small lunar rotation inclination means that the maximum scattering illumination occurs during the lunar mid-summer. The first lunar south pole summer after the SELENE launch in September 2007 was during October to December 2007. Clear images of inside of Shackleton Crater were first acquired on 19 November 2007 (Fig. 1, A and B), from which we also produced a digital terrain model (DTM) (Fig. 1, C and D). Shackleton Crater is a truncated cone-shaped crater and has an almost concentric circular rim with a radius of ~10.5 km, a floor with a radius of ~3.3 km, and a depth of ~4.2 km; it is much deeper than other similarly sized lunar craters (*8*). Inside, the crater has an almost smooth but cratered inner wall. Two mounds are seen adjacent to the inner wall that are probably the result of landslides from the inner wall. A hill a few hundred meters in height occupies the crater center. A terrace structure is associated with the hill and is elongated toward the inner wall. It has a number of depressions that are probably craters.

The inner wall slopes ~30°, which is consistent with a result from the Earth-based radar observation of the upper portion of the Shackleton Crater inner wall that is opposite the Earth (*6*). The entire rim of Shackleton Crater is tilted ~1.5° toward a direction of 50° to ~90° from the Earth-side hemisphere. Thus, under the illumination conditions of the lunar summer, the solar elevation angle from the opposite-side rim occasionally becomes a maximum ~3.0° on a few days when the sub-solar point is ~70°E in longitude and ~1.5°S in latitude, such as on 19 November 2007. Meanwhile, when the sub-solar point becomes far from ~70°E in longitude, the illuminated areas on the inner wall of Shackleton decreases, and the floor remains in darkness. Similarly bright conditions occurred around 18 December 2007 as anticipated, and we obtained additional clear images of the crater interior then. After December 2007, the solar elevation angle decreased, and the inside of Shackleton was less lit. The next time the bottom of Shackleton Crater will be maximally lit will occur around 7 November 2008. Based on the observed shape parameters from the TC DTM, we were able to estimate the surface temperature of the PSA of the crater floor [see the supporting online material (SOM) text] (Fig. 1E). We assumed that the crater has a Lambert diffusive surface and the same albedo between the inner wall and the floor and that the solar elevation angle is 3.0° on the date of its most illuminated condition (for example, 19 November 2007), as in Fig. 1A. As is seen in Fig. 1E, the highest estimated temperature is ~88 K in the center of the floor. The floor temperature was largely determined by the radiation in the infrared range and not by that in the visible range. Thus, the shape parameters, particularly its depth, predominated because the infrared radiation in the crater rapidly decreases as its area increases. The surface visual albedos and scattering laws were almost negligible in estimating the temperature. The loss rate of any ice by vaporization at 90 K is approximately 10^{-26} to 10^{-27} m/s (*9*). Therefore, any water vapor brought here by comets or meteorites could have been trapped for billions of years.

However, we could not find any conspicuously bright areas in Shackleton Crater. The hemispherical visual albedo around the center of the crater

[1]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa 229-85105, Japan. [2]Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [4]NIT DATA CCS Corporation, Koto-ku, Tokyo 136-0071, Japan. [5]Department of Earth and Space Science, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan. [6]University Museum, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. [7]Department of Aerospace and Astronautics, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. [8]Forum Research, Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan. [9]Taito-ku, Tokyo 111-0035, Japan. [10]Division of Earth and Planetary Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [11]Space Exploration Institute, CP 774, CH-2002 Neuchatel, Switzerland.

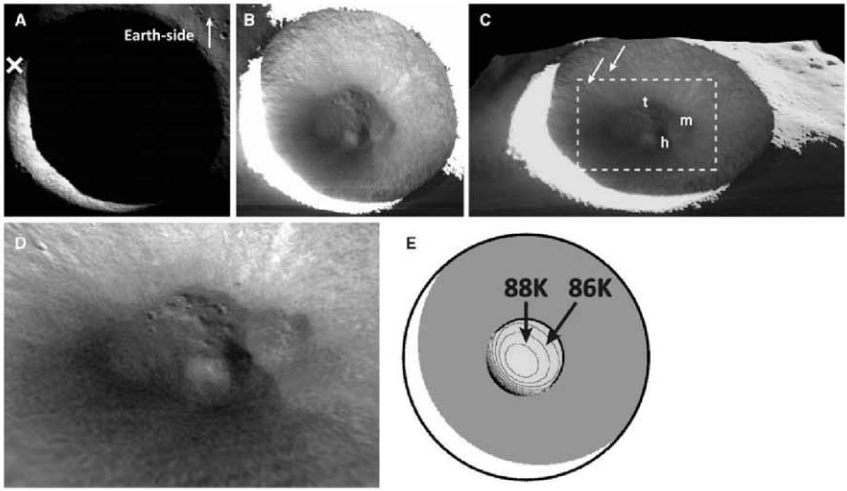


Fig. 1. Inside Shackleton Crater, a crater ~10.5 km across near the lunar south pole. **(A)** Image taken by the SELENE TC on 19 November 2007. The crescent in the image is the portion of the crater's inner wall illuminated by sunlight. The lunar south pole is on the upper left, indicated by cross. **(B)** An enhanced image of **(A)**. The permanently shadowed area inside Shackleton Crater is lit by scattered light from the illuminated portion of the upper wall. No deposit with a significant albedo anomaly is exposed on the floor. **(C)** A perspective view of Shackleton Crater, produced from the DTM that was based on TC stereo-pair data and the brightness-enhanced image [similar to (B)]. A few craters hundreds of meters in diameter exist on the inner wall; some examples are indicated by arrows. The mound-like feature (m) (~300 to 400 m in thickness) is probably the result of downslide movements of materials from the inner wall. The central hill (h) is ~200 m in height and associated with a terrace-like terrain (t) with several hundred-meter–scale craters. **(D)** A closer view of the rectangular area marked in (C). **(E)** Simulated temperature distribution inside the crater floor made with data from 19 November 2007 (see the SOM text for details of the simulation). The contour interval is 2 K. The maximum temperature of the floor is ~88 K near the center, which is cold enough to retain water-ice; however, temperature information alone does not settle whether there is water-ice on the lunar poles. The temperature distribution of the inner wall is not indicated in this figure.

floor was found to be 0.23 ± 0.05 (SOM text), which is similar to that around the crater. The incident angles for the inner wall, which is illuminated by direct solar radiation, and for the floor, which is illuminated by the scattered light from the inner wall, were both ~60°. Under such conditions, the derived hemispherical visual albedo, assuming the Lambert diffusive law, is probably overestimated (*10*). Considering that the lunar far-side averaged albedo is 0.22 (*11*) and that of pure water-ice is ~1.0 (*9*), we conclude that there is no extensively exposed pure water-ice deposit occupying an area larger than that seen in the TC's spatial resolution.

Water-ice on the floor of Shackleton Crater may be "dirty" (mixed with soil and disseminated) at only 1 to ~2 weight percent (*12*). Small amounts of water-ice in a soil mixture do not strongly affect the surface brightness. Because the maximum floor albedo is 0.28, and assuming the albedo of lunar soil to be 0.22 and the albedo of water-ice to be 1.0, the area fraction of the exposed surface ice in an area of TC's resolution (10 by 10 m) is probably less than a few percent [(0.28 – 0.22)/(1 – 0.22)/(a factor due to overestimating the floor albedo)]. Alternatively, water-ice that is present may be buried by a thin regolith (lunar soil) layer and, when exposed by impacts, may be largely removed from the uppermost surface by various space weathering processes (*13*). Another possible explanation of the lunar pole hydrogen is that it is from the direct implantation of solar-wind protons into the lunar surface (*14*). This latter interpretation does not require the presence of any water-ice.

References and Notes

1. W. C. Feldman *et al.*, *Science* **281**, 1496 (1998).
2. K. Watson, B. C. Murray, H. Brown, *J. Geophys. Res.* **66**, 3033 (1961).
3. J. R. Arnold, *J. Geophys. Res.* **84**, 5659 (1979).
4. S. Nozette *et al.*, *Science* **274**, 1495 (1996).
5. S. Nozette *et al.*, *J. Geophys. Res.* **106**, 23253 (2001).
6. D. B. Campbell *et al.*, *Nature* **443**, 835 (2006).
7. J. Haruyama *et al.*, *Earth Planets Space* **60**, 243 (2008).
8. R. J. Pike, in *Impact and Explosion Cratering*, D. J. Roddy, R. O. Pepin, R. B. Merrill, Eds. (Pergamon, New York, 1977) pp. 489–509.
9. T. Mukai *et al.*, *Adv. Space Res.* **19**, 1497 (1997).
10. B. Hapke, in *Theory of Reflectance and Emittance Spectroscopy*, (Cambridge Univ. Press, NY, 1993).
11. NASA, "Lunar surface models NASA space vehicle design criteria" (Technical Rep. SP-8023, NASA, Washington, DC, 1969).
12. W. C. Feldman *et al.*, *J. Geophys. Res.* **105**, 4175 (2000).
13. D. H. Crider, R. R. Vondrak, *J. Geophys. Res.* **108**, 5079 10.1029/2002JE002030 (2003).
14. L. V. Starukhina, *Icarus* **147**, 585 (2000).

15. We thank all the contributors to the SELENE (Kaguya) project, especially members of the project management group (Y. Takizawa, M. Kato, S. Sasaki, R. Nagashima, K. Tsuruda, and H. Mizutani) and the Lunar Imager/Spectrometer (LISM) working group (H. Otake, H. Kawasaki, R. Nakamura, S. Kodama, S. Minami, S. Takechi, A. Akiyama, T. Yokota, T. Arai, T. Sugihara, Y. Yamaguchi, S. Sasaki, N. Asada, H. Demura, N. Hirata, J. Terazono, T. Hirai, T. Hashimoto, T. Michikami, K. Kitazato, M. Higa, P. Pinet, T. Nimura, T. Yamamoto, N. Harada, K. Iseki, T. Hodokuma, S. Kikuchi, S. Kawabe, S. Okuno, and T. Takayama) for their efforts in the development, operation, and data processing of SELENE and LISM/TC. We thank the three anonymous reviewers for their helpful comments. This work was supported by KAKENHI (grants 20540416 to J.H. and C.H. and 20.9211 to T. Morota).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164020/DC1
SOM Text
References

31 July 2008; accepted 30 September 2008
Published online 23 October 2008;
10.1126/science.1164020
Include this information when citing this paper.

A Test of Climate, Sun, and Culture Relationships from an 1810-Year Chinese Cave Record

Pingzhong Zhang,[1] Hai Cheng,[2]* R. Lawrence Edwards,[2] Fahu Chen,[1] Yongjin Wang,[3] Xunlin Yang,[1] Jian Liu,[4] Ming Tan,[5] Xianfeng Wang,[2] Jinghua Liu,[1] Chunlei An,[1] Zhibo Dai,[1] Jing Zhou,[1] Dezhong Zhang,[1] Jihong Jia,[1] Liya Jin,[1] Kathleen R. Johnson[6]

A record from Wanxiang Cave, China, characterizes Asian Monsoon (AM) history over the past 1810 years. The summer monsoon correlates with solar variability, Northern Hemisphere and Chinese temperature, Alpine glacial retreat, and Chinese cultural changes. It was generally strong during Europe's Medieval Warm Period and weak during Europe's Little Ice Age, as well as during the final decades of the Tang, Yuan, and Ming Dynasties, all times that were characterized by popular unrest. It was strong during the first several decades of the Northern Song Dynasty, a period of increased rice cultivation and dramatic population increase. The sign of the correlation between the AM and temperature switches around 1960, suggesting that anthropogenic forcing superseded natural forcing as the major driver of AM changes in the late 20th century.

Interest in Asian Monsoon (AM) variability is high because of tropical ocean/atmosphere interactions involving the AM and the AM's role in transporting large quantities of heat and water vapor to the most-populated regions of the world. Recent cave climate studies have reconstructed the monsoon's history back thousands (*1*) to hundreds of thousands of years (*2*). However, we still do not know, in detail, how late Holocene AM changes relate to solar activity, climate in other regions, anthropogenic forcing, and cultural changes, mainly because of age uncertainties and the insufficient resolution of existing records. We identified an unusual stalagmite sample from Wanxiang Cave, China, that is ideally suited to mitigate these problems. High growth rate, high uranium concentrations, and low thorium concentrations allow high oxygen isotope resolution and high-precision ^{230}Th ages. The cave is located on the fringes of the area currently affected by the summer monsoon and is thus sensitive to and integrates broad changes in the monsoon.

Wanxiang Cave (33°19'N, 105°00'E, 1200 m above sea level) is between the Qinghai-Tibetan Plateau and the Chinese Loess Plateau in Wudu County, Gansu Province, China (fig. S1). Cave and mean annual temperatures are ~11°C. The region is semi-arid (fig. S1), with annual precipitation of 480 mm, 80% of which falls between May and September. During winter, the Siberian-Mongolian High and westerly winds maintain cold and dry conditions and carry airborne dust to the site. In May 2003, about 1 km from the cave entrance, we collected a 118-mm-long stalagmite (WX42B), which had grown continuously from 190 to 2003 A.D. (table S1 and fig. S2). The record was established with 703 δ^{18}O analyses with an average resolution of 2.5 years, errors in ^{230}Th age of <5 years (Fig. 1), and errors associated with subsampling position of <15 years [materials and methods in supporting online material (SOM) and table S2]. A Hendy test (*3*) and other lines of reasoning suggest that the calcite was deposited in isotopic equilibrium and that the δ^{18}O signal largely anticorrelates with precipitation (materials and methods in SOM and figs. S3 and S4). Comparisons with existing cave records from China are difficult because of differing oxygen isotope resolution and dating uncertainties (fig. S5). However, a recent precipitation record [(*4*), from historical documents] from Longxi (~150 km to the north) exhibits close similarity to the cave record (Fig. 2A), further supporting the precipitation/cave δ^{18}O relationship.

The Wanxiang record, with a δ^{18}O range of ~1.3 per mil (‰) (Fig. 1), exhibits a series of centennial to multicentennial fluctuations broadly similar to those documented in Northern Hemisphere (NH) temperature reconstructions, including the Current Warm Period (CWP), Little Ice Age (LIA), Medieval Warm Period (MWP), and Dark Age Cold Period (DACP) (*5*–*8*). Between 190 and 530 A.D. (the end of the Han Dynasty and most of the Era of Disunity), the summer monsoon is moderately strong, but with substantial decadal- to centennial-scale fluctuation. From 530 to 850 A.D. (the end of the Era of Disunity, the Sui Dynasty, and most of the Tang Dynasty), the AM generally declines, then decreases sharply to a relative minimum at 860 A.D. It maintains low values until another sharp drop between 910 and 930 A.D., then rises dramatically over 60 years to a maximum at around 980 A.D., maintaining high values until 1020 A.D. We term the weak monsoon period between 850 and 940 A.D. the Late Tang Weak Monsoon Period (LTWMP; the last six decades of the Tang Dynasty and the first three decades of the Five Dynasties and Ten Kingdoms period). Similarly, we refer to the ensuing strong monsoon peak (960 to 1020 A.D.) as the Northern Song Strong Monsoon Period (NSSMP; the first six decades of the Northern Song Dynasty). After 1020 A.D., the monsoon fluctuates considerably but is generally strong until a sharp drop between 1340 and 1360 A.D.: the mid–14th-century monsoon weakening. It maintains weak values, with substantial fluctuation, until a sharp increase between 1850 and 1880 A.D.: the late–19th-century monsoon strengthening. Punctuating the generally weak monsoon of the 14th to 19th centuries are unusually weak periods from 1350 to 1380 A.D. (the Late Yuan Weak Monsoon Period, LYWMP) and from 1580 to 1640 A.D. (the Late Ming Weak Monsoon Period, LMWMP). The last major feature of the monsoon is a distinct weakening at the end of the 20th century (1960 to 2000 A.D.).

A 9th-century dry period has been invoked as contributing to the decline of the Tang Dynasty (*9*) and the Mayans in Mesoamerica (*10*–*12*), an idea that has generated substantial discussion (*13*). The timing of the LTWMP is consistent with this idea, particularly as the Tang was in decline in its final decades. Because the LTWMP continued beyond the Tang, it may have also contributed to the lack of unity during the Five Dynasties and Ten Kingdoms period. Moreover, the ensuing NSSMP may have contributed to the rapid increase in rice cultivation, the dramatic increase in population, and the general stability at the beginning of the Northern Song Dynasty (*14*). Furthermore, the end of the Yuan and the end of the Ming are both characterized by unusually weak summer monsoons (the LYWMP and the LMWMP), suggesting a link between climate and the demise of these dynasties as well. Whereas other factors would certainly have affected these chapters of Chinese cultural history, our correlations suggest that climate played a key role.

The relationship between the AM and the Swiss record of Alpine glaciation (*8*), portions of which are dated to the year, is clear, with times of weak summer AM correlating with Alpine advances (Fig. 2B). The Swiss record captures a 9th-century glacial advance (that correlates with the LTWMP), followed by glacial retreat to high elevations for much of the late 9th to early 14th centuries (corresponding broadly to the MWP and the generally strong AM between 950 and 1340 A.D.). A prominent advance correlates with the mid–14th-century monsoon weakening and the LYWMP, the first of three LIA advances that correlate with a generally weak AM. The second and largest, centered at about 1600 A.D., corre-

[1]Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environment Sciences, Lanzhou University, Lanzhou 730000, China. [2]Department of Geology and Geophysics, University of Minnesota, Minneapolis, MN 55455, USA. [3]College of Geography Science, Nanjing Normal University, Nanjing 210097, China. [4]Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China. [5]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China. [6]Department of Earth System Science, University of California, Irvine, Irvine, CA 92697, USA.

*To whom correspondences should be addressed. E-mail: cheng021@umn.edu

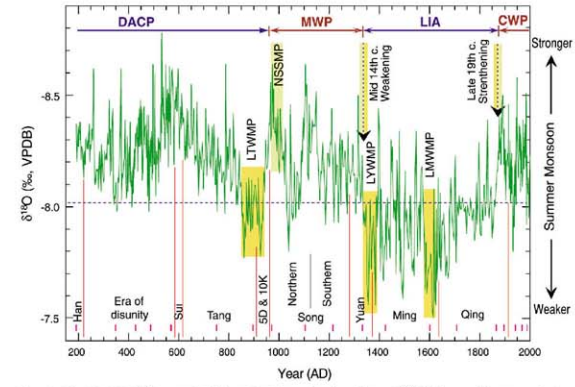


Fig. 1. The WX42B $\delta^{18}O$ record. Pink vertical bars show locations of ^{230}Th dates, with errors of ±1 to ±5 years. The three yellow vertical bars denote the LTWMP, LYWMP, and LMWMP; the shaded green bar denotes the NSSMP. Chinese dynasties are indicated, as are the mid–14th-century monsoon weakening and the late–19th-century monsoon strengthening. 5D & 10K stands for the Five Dynasties and Ten Kingdoms period.

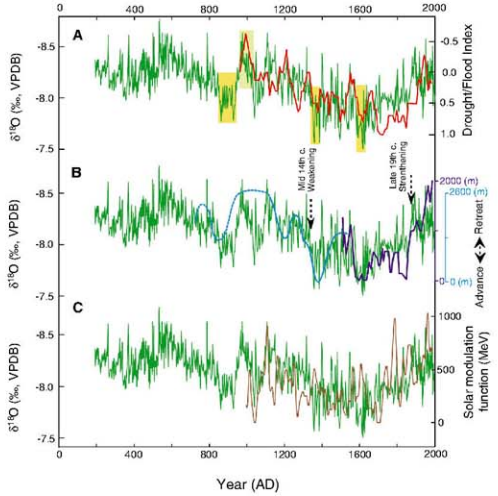


Fig. 2. Comparisons among WX42B, the Longxi drought/flood index, Alpine glacial records, and solar irradiance. The $\delta^{18}O$ time series of WX42B is in dark green. Yellow vertical bars are as in Fig. 1. **(A)** Reconstructed drought/flood index from Longxi, ~150 km north of Wanxiang Cave [red (*4*)]. **(B)** Swiss Alpine glaciation (dark blue, Gorner glacier; light blue, Lower Grindelwald glacier) (*8*). Portions are dated to the year; the dashed portion is dated less precisely (with ^{14}C dates). **(C)** Solar irradiance from ^{10}Be and ^{14}C records [brown (*22*)]. See also fig. S8.

lates with the LMWMP. A retreat, marking the end of the LIA, correlates well with the late–19th-century strengthening of the AM, an event that also correlates with glacial retreat in many localities worldwide and with abrupt freshening and cooling of surface waters in the southwest tropical Pacific (*15*). These correlations demonstrate that the link between North Atlantic/European and East Asian climate, established deeper in time (*1*, *16*), continues into the latest Holocene (except for the end of the 20th century).

The geographic imprint of the LTWMP extends well beyond the Alps, probably including a large portion of the NH low and mid-latitudes. In addition to the aforementioned dry events recorded in Lake Huguang Maar (*9*), Mesoamerica (*10*–*12*), and the Cariaco (*17*), as well as the Alpine glacial advance (*8*), this event correlates with dry events recorded on Barbados (*18*), in southern France (*19*), and in southeastern Minnesota (*20*). The latter two examples correlate with the LTWMP at very high temporal precision. The LTWMP's broad NH footprint supports the idea of a southerly shift in the Intertropical Convergence Zone (*10*), with teleconnections to higher northern latitude sites (fig. S6).

Based on speleothem records with chronologies tuned to ^{14}C records (*1*, *21*), solar forcing has been identified as one factor affecting the AM. Our record provides a rigorous test of this idea (tables S1 and S2). Spectral analysis yields significant periods at 170, 10.5, and 5 to 6 years, with the 10.5-year period corresponding to the Schwabe sunspot cycle (fig. S7). The AM also has correlations to solar irradiance as inferred from ^{14}C and ^{10}Be records (*22*) [correlation coefficient (r) = –0.33, n = 345 data points for the past millennium, Fig. 2C and fig. S8]. These observations support the idea that solar forcing played a role in driving AM changes during the past two millennia (*1*, *5*–*7*, *21*). However, it is clear from this and other studies (*1*, *5*–*7*, *21*) that other factors control much of the variability in the AM. For example, one of the most prominent features in our record, the mid–14th-century monsoon weakening, takes place the better part of a century after a major inferred drop in solar intensity (*22*) (Fig. 2C) and is therefore not likely to be the result of that solar shift.

Summer AM precipitation is driven by the thermal contrast between Asia and the tropical Indo-Pacific and therefore should be linked to regional temperature change (*16*). Based on an ECHO-G climate model simulation (SOM), we modeled Chinese air temperature for the past millennium, with results that were broadly consistent with our record (r = –0.46 and n = 345 data points), particularly with regard to centennial and multicentennial time scales and with regard to the distinct double peak in both model output and record during the MWP (Fig. 3A). Because the ECHO-G model includes solar (and volcanic) forcing, this broad agreement further substantiates the view that solar forcing plays a role in AM variability.

Our record has substantial similarities to NH temperature reconstructions (*5*–*7*) (Fig. 3), with notable differences: (i) The AM peak at 550 A.D. is higher than correlative temperature anomalies (Fig. 3, C and D), and (ii) the timing of the onset (950 A.D.) and end (1340 A.D.) of the strong AM at about the time of the MWP differ from those of the NH temperature anomaly [~890 to 1170 A.D. (*23*)]. In both cases, our record follows solar variability more closely than NH temperature. Solar irradiance reached its highest value in the past two millennia (fig. S8) at the same time as the 550 A.D. AM peak. At about the time of the MWP, the AM and solar irradiance peaks (*24*) have similar timing. These observations suggest that the AM and Chinese temperature respond more strongly to solar irradiance than does mean NH temperature.

Our record shows general AM weakening over the past half century, with much of the change taking place in the past two decades, which is consistent with other AM indices (fig. S9) and anticorrelating with rising NH temperature (*5*–*7*). This late-20th-century anticorrelation is distinctly anomalous as compared to earlier times, which are characterized by millennial- and centennial-scale correlation between NH temperature and the AM (Fig. 2), as well as similar correlations between the AM and North Atlantic temperature much deeper in time (*2*, *16*). This anomaly suggests that the dominant forcing of AM variability changed from natural to anthropogenic around 1960 (*25*–*27*). Possible mechanisms include differences in the nature of solar versus greenhouse forcing, the effect of anthropogenic black carbon on the AM, and/or the effect of anthropogenic sulfate aerosols. Solar forcing is more seasonally and spatially heterogeneous than greenhouse gas forcing and thus is more likely to induce feedbacks involving temperature gradient–driven circulation such as the AM (*25*). Models have also shown that lower-troposphere cooling coupled with mid- and upper-troposphere heating caused by black carbon could reduce East Asian Monsoon precipitation as observed (*28*). Finally, models also suggest that the indirect effect of differential anthropogenic sulfate aerosol loading shifts tropical rainfall southward (*29*), plausibly weakening the monsoon as observed. Thus, variability in late–20th-century precipitation and temperature in the region is probably caused, in large part, by both anthropogenic greenhouse gases and aerosols. In reaching this conclusion, the key observation was not our late–20th-century trend, which is also observed in instrumental records, but rather the relationships at earlier times, observed in this and other proxy records. The proxy records establish the character of natural climate change, from which we distinguish late–20th-century trends as clearly anomalous.

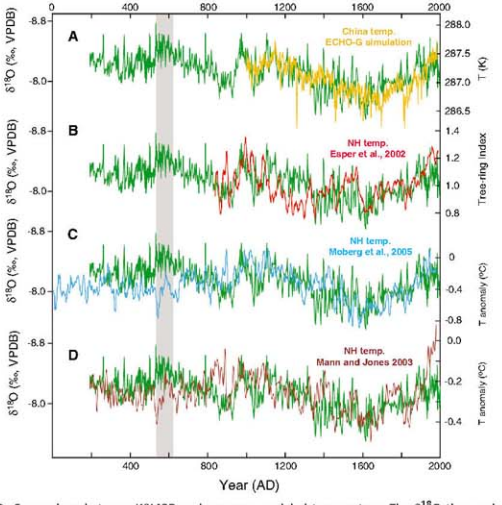


Fig. 3. Comparison between WX42B and proxy or modeled temperature. The $\delta^{18}O$ time series of WX42B is in dark green. **(A)** Modeled Chinese temperature from an ECHO-G model simulation (orange). **(B** to **D)** NH temperature reconstructions from proxy records: (B) red curve (*5*), (C) blue curve [smoothed with a 10-year running mean (*7*)], and (D) dark brown curve (*6*). The gray vertical bar depicts a time characterized by a discrepancy between the AM record and NH temperature reconstructions, as discussed in the text.

Reference and Notes

1. Y. J. Wang *et al.*, *Science* **308**, 1090 (2005).
2. Y. J. Wang *et al.*, *Nature* **451**, 854 (2008).
3. C. H. Hendy, *Geochim. Cosmochim. Acta* **35**, 801 (1971).
4. L. C. Tan, Y. J. Cai, L. Yi, Z. S. An, L. Ai, *Clim. Past* **4**, 19 (2008).
5. J. Esper, E. R. Cook, F. H. Schweingruber, *Science* **295**, 2250 (2002).
6. M. E. Mann, P. D. Jones, *Geophys. Res. Lett.* **30**, 1820 (2003).
7. A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko, W. Karlén, *Nature* **433**, 613 (2005).
8. H. Holzhauser, M. Magny, H. J. Zumbuhl, *Holocene* **15**, 789 (2005).
9. G. Yancheva *et al.*, *Nature* **445**, 74 (2007).
10. J. W. Webster *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **250**, 1 (2007).
11. D. A. Hodell, M. Brenner, J. H. Curtis, T. Guilderson, *Science* **292**, 1367 (2001).
12. J. H. Curtis, D. A. Hodell, M. Brenner, *Quat. Res.* **46**, 37 (1996).
13. D. E. Zhang, L. H. Lu, *Nature* **450**, 10.1038/nature06338 (2007).
14. J. A. G. Roberts, *A History of China* (Palgrave Macmillan, ed. 2, New York, 2006).
15. E. J. Hendy *et al.*, *Science* **295**, 1511 (2002).
16. H. Cheng *et al.*, *Geology* **34**, 217 (2006).
17. G. H. Haug, K. A. Hughen, D. M. Sigman, L. C. Peterson, U. Rohl, *Science* **293**, 1304 (2001).
18. J. L. Banner, M. Musgrove, Y. Asmerom, R. L. Edwards, J. A. Hoff, *Geology* **24**, 1049 (1996).
19. E. M. McMillan, I. J. Fairchild, S. Frisia, A. Borsato, F. McDermott, *J. Quat. Sci.* **20**, 423 (2005).
20. S. Dasgupta, thesis, Univ. of Minnesota, Minneapolis, MN (2008).
21. U. Neff *et al.*, *Nature* **411**, 290 (2001).
22. R. Muscheler *et al.*, *Quat. Sci. Rev.* **26**, 82 (2007).
23. T. J. Osborn, K. R. Briffa, *Science* **311**, 841 (2006).
24. E. Bard, M. Frank, *Earth Planet. Sci. Lett.* **248**, 1 (2006).
25. G. A. Meehl, W. M. Washington, T. M. L. Wigley, J. M. Arblaster, A. Dai, *J. Clim.* **16**, 426 (2003).
26. C. M. Ammann, F. Joos, D. S. Schimel, B. L. Otto-Bliesner, R. A. Tomas, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3713 (2007).
27. M. Lockwood, C. Fröhlich, *Proc. R. Soc. London Ser. A* **463**, 2086 (2007).
28. K. M. Lau, M. K. Kim, K. M. Kim, *Clim. Dyn.* **26**, 855 (2006).
29. L. D. Rotstayn, U. Lohmann, *J. Clim.* **15**, 2103 (2002).

30. We thank R. S. Bradley for his comments that improved this manuscript considerably and W. S. Broecker and the late G. Comer for their support of our work. This work was supported by the National Science Foundations of the United States (grant NSF ESH 0502535 to R.L.E. and H.C.) and China (grants NSFC 40421101, 40471137, and 40721061 to P.Z. and 40225007 to Y.J.W. and H.C.), the Gary Comer Science and Education Foundation (grants CC8 and CP41 to R.L.E.), the Ministry of Science and Technology of China (grant 2006CB400503 to M.T.), and the Chinese Academy of Sciences (grant KZCX2-YW-316 to M.T.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/940/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 and S2
References

30 July 2008; accepted 2 October 2008
10.1126/science.1163965

Recycling of Graphite During Himalayan Erosion: A Geological Stabilization of Carbon in the Crust

Valier Galy,[1,2]* Olivier Beyssac,[3] Christian France-Lanord,[2] Timothy Eglinton[1]

At geological time scales, the role of continental erosion in the organic carbon (OC) cycle is determined by the balance between recent OC burial and petrogenic OC oxidation. Evaluating its net effect on the concentration of carbon dioxide and dioxygen in the atmosphere requires the fate of petrogenic OC to be assessed. Here, we report a multiscale (nanometer to micrometer) structural characterization of petrogenic OC in the Himalayan system. We show that graphitic carbon is preserved and buried in marine sediments, while the less graphitized forms are oxidized during fluvial transport. Radiocarbon dating indicates that 30 to 50% of the carbon initially present in the Himalayan rocks is conserved during the erosion cycle. Graphitization during metamorphism thus stabilizes carbon in the crust over geological time scales.

The burial of organic carbon (OC) in marine sediments represents the second largest sink of atmospheric CO_2 after silicate weathering and subsequent carbonate precipitation (*1–4*). Rivers play a crucial role in this process by exporting a large flux of OC from the continents to the oceans (*5–7*), hence generating a long-term C sink. OC exported by rivers is, however, a mix of (i) "recent" OC (OC_{recent}) derived from plant detritus, associated soil organic matter, and autotrophic carbon production by aquatic plants, and (ii) petrogenic OC (OC_{petro}) derived from erosion of carbonaceous rocks [e.g., (*8–11*)]. Burial of the latter is a simple recycling of reduced C and has no effect on the long-term atmospheric CO_2 and O_2 levels. Conversely, its oxidation consumes O_2 from and returns CO_2 to the atmosphere, thereby counteracting the effect of OC_{recent} burial. Addressing the role of continental erosion on the global C cycle thus requires assessing the fate of OC_{petro} during erosion and fluvial transport.

The Himalayan range undergoes an intense physical erosion producing an enormous flux of sediment (~2 billion tons per year) carried by the Ganges-Brahmaputra (G-B) fluvial system to the Bengal Fan turbiditic system [e.g., (*12*)]. Recently, we showed that OC is very efficiently buried in sediments of the Bengal Fan (*13*). This contrasts with most large deltaic systems, including the Amazon (*5–7*). We estimated that Himalayan erosion is responsible for ~10 to 20% of the global burial of OC_{recent}, implying a global impact on the OC cycle. The Himalayan example illustrates the importance of active orogens in the global C cycle and therefore offers an ideal case study to determine the fate of OC_{petro} during continental erosion.

Several methods have been developed to detect and quantify the most refractory pool of C in rocks, river sediments, and marine sediments, often referred to as black carbon [e.g., (*14–17*)]. The selectivity of these methods is, however, based on the refractory character of the C particles, which is not unique to OC_{petro}, but also characterizes some soil OC (charcoals) or anthropic C (e.g., soots). Consequently, they isolate a much larger pool of C and thus overestimate the petrogenic component. Here, we combine Raman microspectroscopy (RM) and transmission electron microscopy (TEM) to (i) detect OC_{petro} in river and marine sediments and (ii) characterize its multiscale structural organization (nm to μm). RM allows in situ analysis, preserving the textural information of OC (particle size and shape) itself, but also its relations with minerals (aggregates and inclusions). Combining 002 lattice-fringes imaging and electron diffraction, TEM gives an insight into the nm-scale structure of the aromatic skeleton composing OC_{petro} (*16*, *17*). Finally, we use radiocarbon dating of bulk acid-insoluble OC to quantify the proportion of OC_{petro} and OC_{recent} in river sediments (*17*).

We have analyzed a sample set covering the whole erosional and depositional system, including Himalayan source rocks, G-B river sediments, and Bengal Fan sediments (tables S1 and S2). To study the fate of OC_{petro} during fluvial transport, river

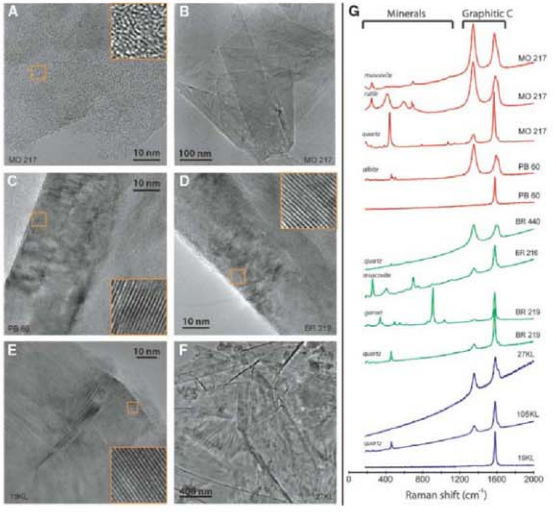


Fig. 1. RM and TEM characterization of OC_{petro} in riverine and marine sediments. **(A)** Image of 002 lattice-fringes (002LF) of disordered and microporous OC_{petro} in Narayani bed load. **(B)** Image (002LF) of graphitic C in Narayani bed load. **(C)** Low-magnification (LM) image of a graphite particle in Narayani suspended load. **(D)** Image (002LF) of graphitic OC_{petro} in Lower Meghna bed load. **(E)** Image (002LF) of graphitic OC_{petro} in distal Bengal Fan sediment. **(F)** LM image of graphite particles in distal Bengal Fan sediment. In (A), (C), (D), and (E), the orange inset is a zoom (5-nm side) located on the main image. **(G)** Selection of representative Raman spectra from Himalayan rivers (red), Lower Meghna (green), and Bengal Fan sediments (blue). These spectra were obtained from both individual OC_{petro} particles (no mineral contribution in the spectrum) and OC_{petro} inclusions/aggregates within minerals (mineral contribution in the spectrum as depicted).

[1]Woods Hole Oceanographic Institution, Department of Marine Chemistry and Geochemistry, Woods Hole, MA 02543, USA. [2]Centre de Recherches Pétrographique et Géochimiques, CNRS, Institut National des Sciences de l'Univers (INSU), Nancy Université, BP 20, 54501 Vandoeuvre-lès-Nancy, France. [3]Laboratoire de Géologie, Ecole Normale Supérieure, CNRS-UMR 8538, 24 Rue Lhomond, F-75005 Paris Cedex 5, France.

*To whom correspondence should be addressed. E-mail: vgaly@whoi.edu

sediments were collected from the Himalayan range to the Bangladesh delta (fig. S1). In the G-B system, suspended and bed sediments define a continuum from fine, clay-rich, surface-suspended sediments to coarse, quartz-rich, bed sediments. OC content increases with decreasing grain size and increasing proportions of phyllosilicates (*18*). To take this variability into account, we analyzed OC in suspended sediments collected along depth profiles within the river as well as from dredged riverbed sediments. To describe the Bengal Fan turbiditic system, we used sediments drilled by the research vessel (RV) Sonne cruise 93 (*19*) in the shelf, the middle fan channel levee system, and the deep fan (fig. S2).

OC_{petro} has been detected by RM and TEM in all investigated sediments, including the most distal regions of the Bengal Fan (Fig. 1). Using RM, OC_{petro} was identified in various forms: (i) discrete "free" particles from a few to several tens of μm size; (ii) inclusions within quartz, calcite, and metamorphic minerals; and (iii) aggregates with minerals, mostly micas. The latter two have an obvious petrogenic origin, whereas in the former the most disordered C may be OC_{recent} (soot or charcoal) or OC_{petro} from very low-grade metamorphic rocks (*17*).

RM shows a large structural variety of OC_{petro} in bed and suspended loads from Himalayan rivers (Fig. 1). This is confirmed by TEM investigations that reveal the presence of turbostratic disordered OC_{petro} to perfectly crystalline graphite in these samples. The structural variety of OC_{petro} reflects the zoning of metamorphism in the source rocks of the Himalayan range [e.g., (*20*)], from low- to high-grade metamorphism (fig. S1). Similar structural variety is observed within the vertical depth profiles collected at the outflow of the range and in Bangladesh, although the less-ordered OC_{petro} is rarely observed. In all marine sediments, both RM and TEM investigations show that OC_{petro} detected by these methods is almost exclusively highly ordered polycrystalline or pristine graphite (Fig. 1) and that disordered OC_{petro} is rare (*17*). Because we do not observe any segregation in terms of crystallinity of OC_{petro} along vertical depth profiles in rivers, we posit that OC_{petro} is homogenously transported as a continuum through the river section. However, oxidation of the less-graphitic OC_{petro} occurs within the Himalayan range and during the floodplain transit, because these forms are virtually absent in the Bengal Fan sediments, which is consistent with OC_{recent} oxidation and replacement documented in the Gangetic floodplain (*18*). Such selective oxidation might be related to the higher chemical reactivity due to the presence of atoms other than C (primarily H and O) and enhanced by the high nano- to micro-porosity of disordered OC_{petro} compared with graphite.

Estimating the proportion of OC contained in the Himalayan rocks that is ultimately oxidized during the erosion cycle requires quantifying the amount of OC_{petro} in the sediments exported by G-B rivers. We measured the ^{14}C composition of bulk acid-insoluble OC (expressed as the fraction of modern C, Fm) in sediments collected along depth profiles. OC_{petro} is radiocarbon free (Fm = 0), whereas OC_{recent} has variable amounts of radiocarbon (Fm > 0) depending on its residence time in the basin. In the Himalayan basin, a consequence of intense physical erosion is that none of the OC_{recent} particles contained in soils is radiocarbon free, which gives to OC_{petro} a unique geochemical signature (Fm = 0).

The results are presented in Fig. 2 as modern OC content (Fm × %OC) plotted as a function of the total OC content (see also fig. S4 and table S1). Sediments collected along depth profiles define linear trends in this diagram. Suspended and bed sediments from each depth profile thus have identical amount of OC_{petro} (*17*). This consistency derives from the generation of sediment continuums by mixing processes during fluvial transport. A statistical analysis of the trends drawn in Fig. 2 allows the determination of the amount of OC_{petro} (*17*). Large trans-Himalayan rivers sampled at the outflow of the range appear to carry variable amounts of OC_{petro}. We estimate an OC_{petro} content of 0.05% and 0.03% for the Narayani and Kosi, respectively. This variability likely derives from intrinsic characteristics of the

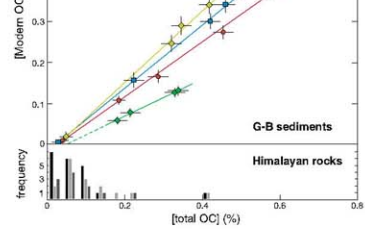


Fig. 2. (Top) Modern OC content (Fm × %OC) of sediments collected over depth profiles in rivers from the G-B system as a function of their total OC content. In this representation, sediments with similar amount of OC_{petro} define linear trends, whose intercept with the *x* axis gives $\%OC_{petro}$ (*17*). Each solid line represents the best linear fit defined by sediments from the same depth profile. Himalayan rivers sediments (green, Narayani; yellow, Kosi) have variable amounts of OC_{petro}, whereas Ganges (blue) and Brahmaputra (red) in Bangladesh have similar amounts of OC_{petro} around an average value of 0.025%. Radiocarbon composition of Bengal Fan sediments (hexagons) is comparable with that of river sediments, indicating they have similar content of OC_{petro}. **(Bottom)** Total OC content of Himalayan source rocks (black, High-Himalayan Crystalline; light gray, Lesser Himalaya; dark gray, composite gravel samples) from (*13*). The average OC content of source rocks ranges between 0.05 and 0.08%.

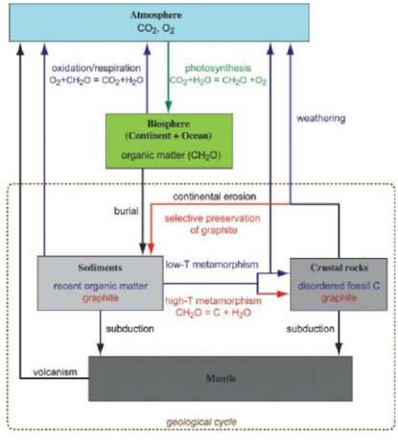


Fig. 3. A model of the long-term OC cycle highlighting the role of graphitization. The graphite formed during high-temperature (>500°C) metamorphism is preserved during continental erosion and returns to the atmosphere only throughout volcanism. Graphitization thus stabilizes reduced C, confining it to a geological subcycle and maintaining an imbalance between photosynthesis and respiration/oxidation over geological time scales.

drainage basins. Our data consistently indicate an OC_{petro} content of 0.02 to 0.03% in the sediments exported by the Brahmaputra and the Ganges sampled close to their mouths in the Bangladesh delta. We therefore estimate that sediments delivered to the Bay of Bengal contain ~0.025% of OC_{petro} derived from the erosion of Himalayan rocks. Because the oxidation of OC in the marine system before deposition in the Bengal Fan is negligible (*13*), the OC_{petro} content of these sediments must be comparable to that of the sediments exported by G-B rivers. Moreover, for two sediments deposited in the channel-levee system of the Bengal Fan, we corrected the ^{14}C composition of OC for their respective deposition age determined by ^{14}C dating of foraminifera (*21*). Their position in Fig. 2 is highly consistent with sediments delivered by G-B rivers, indicating that they have similar content of OC_{petro}. Consequently, we estimate that Bengal Fan sediments contain ~0.025% of OC_{petro}. To evaluate the fate of petrogenic OC during the Himalayan erosion cycle, this value must be compared to the mean OC content of Himalayan rocks. The latter has been estimated to be 0.05 to 0.08%, on the basis of individual rock samples and composite gravels extracted from the bed of Himalayan rivers (*13*, *22*, *23*). Based on these figures, at least 30% (40 ± 10) of the OC contained in the Himalayan rocks appears to be preserved and recycled during the erosion cycle, mostly through burial of highly graphitic carbon in the Bengal Fan sediments. In turn, the oxidative flux of OC_{petro} associated with Himalayan erosion is estimated to be 0.67 (±0.25) × 10^{11} mol/year, which is comparable to the long-term CO_2 consumption by silicate weathering (0.6 × 10^{11} mol/year (*24*)) but one order of magnitude lower than the CO_2 consumption by OC_{recent} burial in the Bengal Fan [3.1 (±0.3) × 10^{11} mol/year (*13*)].

The Himalayan example highlights the role of mountain building in the long-term stabilization of Earth climate and atmospheric composition. Active orogenesis not only promotes CO_2 consumption by both silicate weathering and terrestrial organic carbon burial but also stabilizes C in the crust over geological time scales through graphitization of OC (Fig. 3). Metamorphism during orogenesis actually transforms labile photosynthetic C into refractory petrogenic C through graphitization, the efficiency of this transformation being mainly controlled by the metamorphic temperature. The selective preservation of graphite during erosion and weathering indicates that graphitization occurring during metamorphism subtracts C from the external cycle (atmosphere-biosphere-ocean) and locks it into the geological cycle (Fig. 3). Mechanisms capable of returning graphitic C into the atmosphere involve subduction and volcanism, both having a very long characteristic time compared with surface processes. At geological time scales, the graphitization thus maintains an imbalance between photosynthesis and respiration by locking reduced C into the crust, hence promoting both consumption of CO_2 and accumulation of O_2 in the atmosphere.

References and Notes

1. R. A. Berner, *Science* **249**, 1382 (1990).
2. J. M. Hayes, J. R. Waldbauer, *Phil. Trans. R. Soc. London B Biol. Sci.* **361**, 931 (2006).
3. R. M. Garrels, A. Lerman, F. T. Mackenzie, *Am. Sci.* **64**, 306 (1976).
4. A. C. Lasaga, R. A. Berner, R. M. Garrels, in *The Carbon Cycle and Atmospheric CO2: Natural Variations Archean to Present*, E. T. Sundquist, W. S. Broecker, Eds. (American Geophysical Union, Washington, 1985), vol. 32, pp. 397–411.
5. D. J. Burdige, *Global Biogeochem. Cycles* **19**, GB4011 (2005).
6. J. I. Hedges, R. G. Keil, R. Benner, *Org. Geochem.* **27**, 195 (1997).
7. B. Schlünz, R. R. Schneider, *Int. J. Earth Sci.* **88**, 599 (2000).
8. N. E. Blair *et al.*, *Geochim. Cosmochim. Acta* **67**, 63 (2003).
9. W. M. Sackett, C. W. Poag, B. J. Eadie, *Science* **185**, 1045 (1974).
10. N. E. Blair, E. L. Leithold, R. C. Aller, *Mar. Chem.* **92**, 141 (2004).
11. A. F. Dickens, Y. Gélinas, C. A. Masiello, S. Wakeham, J. I. Hedges, *Nature* **427**, 336 (2004).
12. A. Galy, C. France-Lanord, *Geology* **29**, 23 (2001).
13. V. Galy *et al.*, *Nature* **450**, 407 (2007).
14. Y. Gelinas, K. M. Prentice, J. A. Baldock, J. I. Hedges, *Environ. Sci. Technol.* **35**, 3519 (2001).
15. P. R. Haberstroh *et al.*, *Geochim. Cosmochim. Acta* **70**, 1483 (2006).
16. C. A. Masiello, *Mar. Chem.* **92**, 201 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. V. Galy, C. France-Lanord, B. Lartiges, *Geochim. Cosmochim. Acta* **72**, 1767 (2008).
19. V. Ittekkot, H. R. Kudrass, D. Quadfasel, D. Unger, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **50**, 853 (2003).
20. O. Beyssac, L. Bollinger, J.-P. Avouac, B. Goffe, *Earth Planet. Sci. Lett.* **225**, 233 (2004).
21. V. Galy *et al.*, *Quat. Sci. Rev.* **27**, 1396 (2008).
22. A.-M. Aucour, C. France-Lanord, K. Pedoja, A.-C. Pierson-Wickmann, S. M. F. Sheppard, *Global Biogeochem. Cycles* **20**, GB2006 (2006).
23. C. France-Lanord, L. A. Derry, *Nature* **390**, 65 (1997).
24. A. Galy, C. France-Lanord, *Chem. Geol.* **159**, 31 (1999).

25. We thank H. Kudrass, who provided Bengal Fan samples collected during the RV Sonne cruise; M. Rahman from Dhaka University; A. Gajurel from Tribhuvan University (Kathmandu) for their help during field work in Bangladesh and Nepal; and A. Galy for insightful comments, as well as J. Hayes and an anonymous referee for constructive reviews. This study was supported by the French INSU program "Reliefs de la Terre" and a scholarship from Woods Hole Oceanographic Institution to V.G.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/943/DC1
Material and Methods
SOM Text
Figs. S1 to S4
Tables S1 and S2
References

5 June 2008; accepted 2 October 2008
10.1126/science.1161408

Induced Pluripotent Stem Cells Generated Without Viral Integration

Matthias Stadtfeld,[1,2,4,5] Masaki Nagaya,[3,5] Jochen Utikal,[1,2,4,5] Gordon Weir,[3,5] Konrad Hochedlinger[1,2,4,5]*

Pluripotent stem cells have been generated from mouse and human somatic cells by viral expression of the transcription factors Oct4, Sox2, Klf4, and c-Myc. A major limitation of this technology is the use of potentially harmful genome-integrating viruses. We generated mouse induced pluripotent stem (iPS) cells from fibroblasts and liver cells by using nonintegrating adenoviruses transiently expressing Oct4, Sox2, Klf4, and c-Myc. These adenoviral iPS (adeno-iPS) cells show DNA demethylation characteristic of reprogrammed cells, express endogenous pluripotency genes, form teratomas, and contribute to multiple tissues, including the germ line, in chimeric mice. Our results provide strong evidence that insertional mutagenesis is not required for in vitro reprogramming. Adenoviral reprogramming may provide an improved method for generating and studying patient-specific stem cells and for comparing embryonic stem cells and iPS cells.

The introduction of defined transcription factors into mouse and human somatic cells has recently been shown to reprogram the developmental state of mature cells into that of pluripotent embryonic cells, generating so-called "induced pluripotent stem (iPS) cells" (*1*). iPS cells have been generated from multiple cell types by viral expression of Oct4 and Sox2, combined with either Klf4 and c-Myc (*1*–*11*) or LIN28 and Nanog (*12*). iPS cells are molecularly and functionally highly similar to ES cells, which makes in vitro reprogramming an attractive approach to produce patient-specific stem cells for studying and potentially treating degenerative disease. Indeed, reprogrammed skin cells have recently been shown to alleviate the symptoms of Parkinson's disease (*13*) and sickle cell anemia (*14*) in mouse models. However, a major limitation of this technology is the use of viruses that integrate into the genome and are associated with the risk of tumor formation due to the spontaneous reactivation of the viral transgenes (*8*). The low efficiency of reprogramming (0.01 to 0.1% of input cells) also raised the possibility that insertional mutagenesis may be a prerequisite for in vitro reprogramming (*15*). For example, the retroviral tagging of explanted hematopoietic stem cells has been previously shown to select for clones in which the retroviral construct had inserted proximal to self-renewal genes and thus causes their activation (*16*). Whereas the sequencing of a limited number of insertion sites in iPS cells did not reveal common targets (*17*), this possibility has not been unequivocally ruled out yet (*15*).

Therefore, we set out to generate iPS cells from mouse somatic cells using adenoviral vectors that allow for transient, high-level expression of exogenous genes without integrating into the host genome (*18*). Specifically, we cloned the cDNAs for *Oct4*, *Sox2*, *c-Myc*, and *Klf4* into replication-incompetent adenoviral vectors under the control of the human cytomegalovirus immediate early (hCMV IE) promoter (fig. S1). Initial attempts to reprogram mouse tail-tip fibroblasts (TTFs) into iPS cells with adenoviruses expressing the four reprogramming factors were unsuccessful, possibly owing to the rapid dilution of the virally encoded proteins from the cells. We have previously shown that viral Oct4 expression can be replaced by a doxycycline-inducible *Oct4* allele driven by a reverse-tetracycline–dependent transactivator (rtTA) present in the ROSA26 locus ($Oct4^{IND}$) (*6*, *19*), and it has been reported that liver cells require lower numbers of viral integrations than fibroblasts to be reprogrammed (*17*). Therefore, we infected ~500,000 adherent $Oct4^{IND}$ mouse fetal liver cells with adenoviruses expressing Sox2, Klf4, and c-Myc at multiplicities of infection (MOI, number of viral particles per cell) of 20 to 50. This led to an infection efficiency of about 40 to 50% of cells with each factor and an estimated infection efficiency of 20 to 30% of cells expressing all three adenoviral transgenes (fig. S2, A and E). After culture of infected fetal liver cells for 24 to 30 days in the presence of doxycycline, nine iPS-like colonies emerged that expressed the pluripotency markers Sox2 and SSEA-1 and could be expanded into stable ES cell–like lines, similar to iPS cells produced with retro- or lentiviral vectors (*5*, *20*) (Fig. 1A). These adenoviral iPS (adeno-iPS) cells continued to grow in the absence of doxycycline, which indicated that transgenic *Oct4* expression was no longer required.

These results prompted us to test if postnatal tail fibroblasts carrying the *Oct4*-inducible allele were equally amenable to reprogramming into adeno-iPS cells. Fibroblasts required MOIs of 50 to 250 to achieve an infection efficiency of ~30% for each vector, resulting in an estimated 10 to 20% infection efficiency for triple-infected cells (fig. S2, B and E). Infection of more than 1,000,000 neonatal $Oct4^{IND}$ TTFs harboring an *Oct4-GFP* reporter with adenoviruses expressing *myc*, *Klf4*, and *Sox2* in the presence of doxycycline gave rise to a single GFP^+ colony that grew into a stable, doxycycline-independent line.

To identify an adult cell type that might not require transgenic Oct4 expression, we chose hepatocytes, which are highly permissive for adenoviral infection (*20*, *21*). Indeed, MOIs of 1 to 4 were sufficient to infect 70 to 80% of these cells with individual vectors, with an estimated 50 to 60% of cells expressing all four viral reprogramming factors (fig. S2, C, D, and E). After incubation of ~500,000 adult hepatocytes isolated

[1]Massachusetts General Hospital Cancer Center and Center for Regenerative Medicine, 185 Cambridge Street, Boston, MA 02114, USA. [2]Harvard Stem Cell Institute, Cambridge, MA 02138, USA. [3]Section on Islet Transplantation and Cell Biology, Joslin Diabetes Center, One Joslin Place, Boston, MA 02115, USA. [4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [5]Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: khochedlinger@helix.mgh.harvard.edu

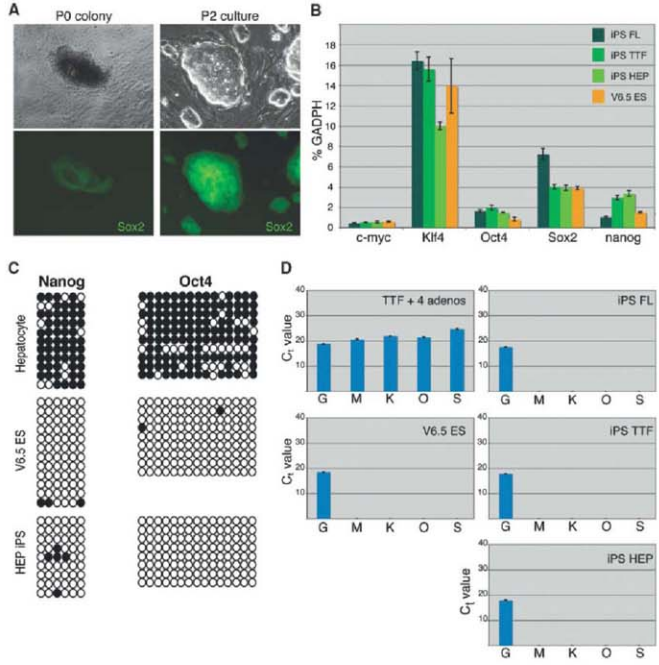


Fig. 1. Analysis of pluripotency markers in adeno-iPS cells. **(A)** Bright-field (top) and fluorescence (bottom) images of an adeno-iPS cell clone established from Sox2-GFP fetal liver cells taken at passage 0 (P0) and passage 2 (P2). **(B)** Expression of endogenous *c-myc*, *Klf4*, *Oct4*, *Sox2* and *nanog* measured by qPCR in adeno-iPS cells derived from fetal liver (FL), fibroblasts (TTF), and hepatocytes (HEP), as well as in V6.5 control ES cells. **(C)** Bisulfite sequencing of the *Oct4* and *Nanog* promoters in hepatocytes, ES cells, and iPS cells derived from hepatocytes. Open circles represent unmethylated CpGs; filled circles denote methylated CpGs. **(D)** Expression levels of endogenous *gapdh* (G), as well as adenoviral *c-myc* (M), *Klf4* (K), *Oct4* (O), and *Sox2* (S) in fibroblasts 3 days after infection with adenoviruses (TTF + 4 adenos), ES cells and adeno-iPS cells derived from fetal liver, fibroblasts, and hepatocytes. Error bars indicate 1 SD. The absence of blue bars in (D) indicates that the respective cDNA was not detected.

from mice carrying the *Oct4-GFP* and *ROSA26-rtTA* alleles but lacking the *Oct4*-inducible allele with adenoviruses expressing c-*myc*, *Klf4*, *Oct4*, and *Sox2*, we obtained three colonies, all of which could be expanded into stable ES-like cell lines expressing the pluripotency markers Oct4 and SSEA-1 (fig. S3). Polymerase chain reaction (PCR) fingerprinting of adeno-iPS cells confirmed their derivation from hepatocytes rather than from a contaminating ES cell line (fig. S4). This demonstrates that iPS-like cells can be produced from adult cells by adenoviral vectors alone in the absence of the transgenic *Oct4* allele.

Next, we tested whether adeno-iPS cells had reestablished pluripotency at the molecular level by examining the activity of ES cell–specific markers. Expression analysis for the endogenous *Oct4*, *Klf4*, *Sox2*, c-*Myc*, and *Nanog* genes by quantitative PCR (qPCR) gave signals that were indistinguishable from those of ES cells, consistent with faithful molecular reprogramming (Fig. 1B). In agreement with this, the *Oct4* and *Nanog* promoters became demethylated in adeno-iPS cells to an extent similar to that seen in ES cells, whereas they remained hypermethylated in cultured fibroblasts and hepatocytes, which indicated that adeno-iPS cells had undergone successful epigenetic reprogramming (Fig. 1C and fig. S5). In contrast to freshly infected fibroblasts, viral transcripts could not be detected in any of the adeno-iPS cell lines tested, which suggested that the viral vectors had been diluted from the cells over time (Fig. 1D).

Adenoviral vectors can integrate into the genome of host cells at extremely low frequencies (*22*). To exclude the possibility of permanent viral integration, PCR analysis of genomic DNA isolated from adeno-iPS cell clones was performed with primers recognizing the different cDNA expression cassettes (see fig. S1 for the position of the sequences recognized by the primers). Whereas adenoviral vector DNA, used as a positive control, readily produced PCR signals, we were unable to amplify these PCR products from genomic DNA from any of the adeno-iPS cells (Fig. 2B). Southern blot analysis using the cDNAs of the four viral vectors as probes confirmed the PCR results and yielded no evidence for the continuous presence of the adenoviral sequences in the adeno-iPS cells whereas the single-copy *Oct4* transgenic allele integrated into the *Col1A* locus could be readily detected in iPS cell clones generated from $Oct4^{IND}$ cells (fig. S6). To rule out genomic integration of adenoviral sequences other than the cDNAs, we performed Southern blot analysis using the Bam HI–digested full-length vector as a probe. Again, we did not detect exogenous viral sequences in the genomes of adeno-iPS cells, whereas adenoviral sequences were detectable in human embryonic kidney (HEK) cells, consistent with a previous report (*23*) (Fig. 2C). Moreover, the pBluescript (pBS)-derived portion of the adenoviral vector probe cross-hybridized with the *Oct4* transgene, which also carries the pBS backbone, giving rise to a specific ~3-kb signal in the iPS cell lines derived from fetal liver and TTFs and thus serving as an internal positive control (Fig. 2C). Together, these results strongly suggest that viral integration is not required for the generation of iPS cells. Although highly unlikely, we cannot rule out the possibility that small pieces of adenoviral DNA have inserted into the genome of adeno-iPS cells but were not observed because of the detection limits of Southern blot analysis.

To ascertain the developmental potential of adeno-iPS cells, we injected the cells into the flanks of SCID mice. All cell lines tested produced teratomas after 3 to 4 weeks, which, upon histological examination, showed differentiation into representative cell types of the three germ

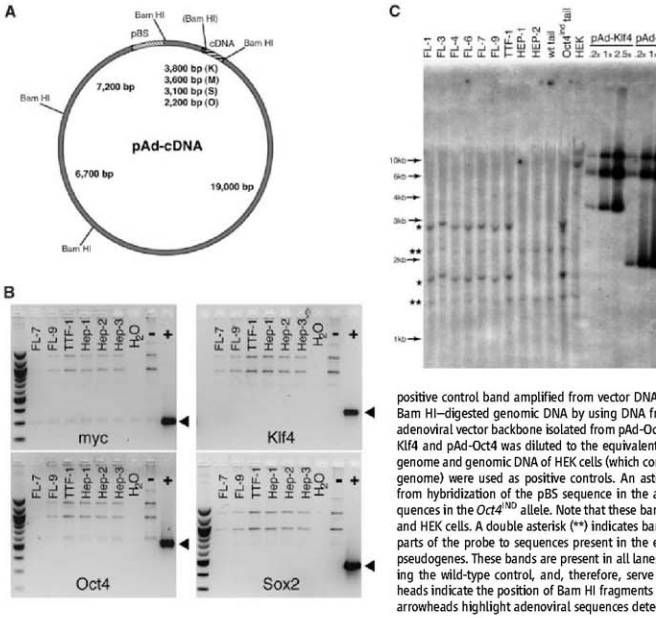


Fig. 2. Absence of viral integration in adeno-iPS cells. **(A)** Schematic drawing of the adenoviral vector indicating the position of the cDNA and the sizes of the respective DNA fragments after Bam HI digestion. The bracketed Bam HI site is only present in the *Oct4* cDNA. A pBluescript (pBS) sequence present in both the adenoviral vector and the $Oct4^{IND}$ transgene is highlighted. **(B)** PCR analyses for adenoviral integration in genomic DNA from the indicated adeno-iPS clones, as well as from V6.5 ES cells that served as a negative control (-). Arrowhead indicates the position of the positive control band amplified from vector DNA (+). **(C)** Southern blot analysis of Bam HI–digested genomic DNA by using DNA fragments encompassing the entire adenoviral vector backbone isolated from pAd-Oct4 as probes. Plasmid DNA of pAd-Klf4 and pAd-Oct4 was diluted to the equivalent of 0.2, 1, or 2.5 integrations per genome and genomic DNA of HEK cells (which contain adenoviral sequences in their genome) were used as positive controls. An asterisk (*) indicates bands resulting from hybridization of the pBS sequence in the adenoviral probe to transgenic sequences in the $Oct4^{IND}$ allele. Note that these bands are absent in HEP iPS, V6.5 ES, and HEK cells. A double asterisk (**) indicates bands resulting from hybridization of parts of the probe to sequences present in the endogenous *Oct4* locus or in *Oct4* pseudogenes. These bands are present in all lanes containing genomic DNA, including the wild-type control, and, therefore, serve as loading controls. Filled arrowheads indicate the position of Bam HI fragments of the adenoviral vector, and open arrowheads highlight adenoviral sequences detected in HEK cells.

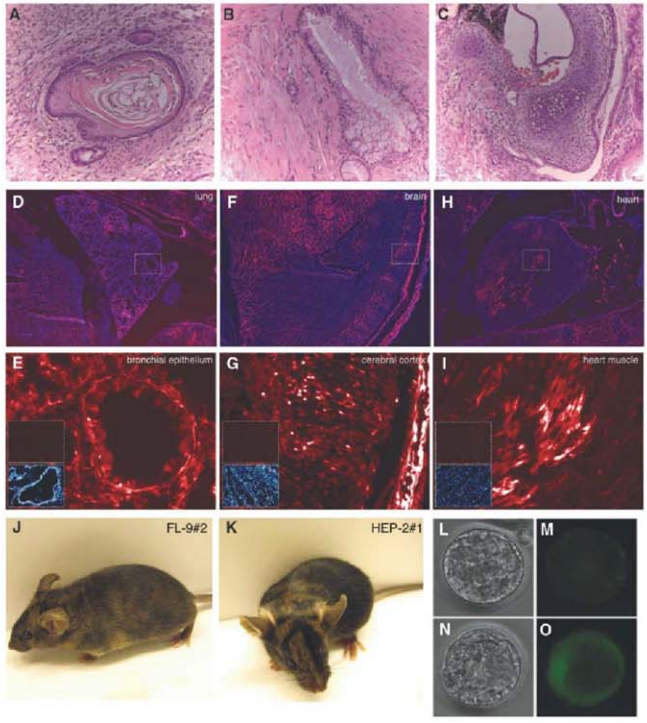


Fig. 3. Pluripotency of adeno-iPS cells. (**A** to **C**) Images of histological sections through teratomas formed by adeno-iPS cells subjected to hematoxylin-and-eosin staining, showing keratinized epithelium (A), mucous epithelium (B) and cartilage (C). (**D** to **I**) Fluorescence images showing the contribution of red fluorescent protein–labeled adeno-iPS cells to lung, brain, and heart in a postnatal chimeric animal. Nuclei were counterstained with DAPI (blue). The small insets in (D), (F), and (H) highlight the fields magnified in (E), (G), and (I); the insets in (E), (G), (I) show the background fluorescent levels and DAPI staining of corresponding tissues in a nonchimeric littermate. (**J** and **K**) Images of coat-color chimeras derived from fetal liver (J) and hepatocytes (K) adeno-iPS cells. (**L** to **O**) Fluorescence and bright-field images of a wild-type (L and M) blastocyst and an *Oct4-GFP* (N, O) blastocyst obtained after mating a chimeric mouse made with TTF-1 iPS cells expressing GFP from the *Oct4* promoter with a wild-type female.

Table 1. Derivation and characterization of adeno-iPS cells from different cell types. $Oct4^{ind}$ indicates $Oct4^{ind}$ allele present; GFP, GFP reporter gene present; ES, expression of ES cell markers (SSEA-1 and Oct4 or Sox2); TF, teratoma formation; PC, postnatal chimeras; GL, germline transmission. ND, not determined.

Source of cells	$Oct4^{ind}$	GFP	Sex	MOI	iPS lines (2n/4n)	Efficiency (iPS/input, %)	ES	TF	PC	GL
Fetal liver	Yes	No	F	20–50	6/0	0.0012	Yes	Yes	ND	ND
Fetal liver	Yes	Sox2	M	20–50	2/1	0.0006	Yes	Yes	Yes	Yes
Fibroblast	Yes	Oct4	M	50–250	1/0	<0.0001	Yes	Yes	Yes	Yes
Hepatocyte	No	Oct4	F	1–4	1/2	0.0005	Yes	Yes	Yes	ND

layers including muscle, cartilage, and epithelial cells, thus demonstrating the pluripotency of adeno-iPS cells (Fig. 3, A to C; also see Table 1 and table S1 for a summary of all adeno-iPS lines). In addition, adeno-iPS cells generated apparently normal postnatal chimeras following injection into blastocysts. One chimera, obtained after blastocyst injection of adeno-iPS cells labeled with a lentivirus expressing the red fluorescent protein tdTomato, was killed at birth to examine the contribution of adeno-iPS cells to different tissues. As shown in Fig. 3, D to I, a high degree of chimerism was seen in multiple tissues including the lungs, brain, and heart. Adeno-iPS cells gave rise to high-degree coat color chimeras (Fig. 3, J and K) and differentiated into functional germ cells, as evidenced by the derivation of GFP^+ blastocysts after breeding of a male chimera generated with TTF-derived adeno-iPS cells with a wild-type female (Fig. 3, L to O). Moreover, adult male chimeras derived from FL-9 and TTF-1 adeno-iPS cells gave rise to viable germline offspring that showed the iPS cell-specific agouti coat color after mating with BDF1 female mice (1/20 or 5% of FL-9 offspring and 11/11 or 100% of TTF-1 offspring). Together, these results indicate that adeno-iPS cells share the same developmental potential as iPS cells obtained with integrating viruses. Importantly, we have not observed tumor formation in any of the 12 coat color chimeras ranging up to 20 weeks of age.

The efficiency of deriving iPS-like cells from fetal liver cells, TTFs, and hepatocytes (see fig. S7 for an illustration of the derivation process from the different cell types) was extremely low, ranging from less than 0.0001% to 0.001% (Table 1). This frequency is lower than that obtained with

integrating viruses (~0.01 to 0.1%) and is probably due to the fact that many cells do not maintain viral expression long enough to trigger entry into a state sustained by endogenous pluripotency factors (*24*, *25*). This conclusion is supported by qPCR analysis for adenoviral gene expression, which is gradually lost in dividing fibroblasts (fig. S8). It should be informative to test whether the low efficiency of adenoviral reprogramming can be increased by the use of chemical compounds as has been reported for retroviral reprogramming (*26*–*28*).

DNA content analysis showed that 3 out of 13 (or about 23%) of the 13 adeno-iPS lines were tetraploid, which is not seen in iPS cells produced with retro- or lentiviral vectors (fig. S9 and Table 1). We speculate that adenoviral reprogramming either induces cell fusion or, alternatively, selects for rare tetraploid cells pre-existing in the starting cell populations. Indeed, it has been shown that the frequency of polyploid hepatocytes increases with age (*29*).

Our results demonstrate the generation of iPS cells without the use of integrating viruses by employing either a combination of adenoviruses and an inducible transgene or adenoviruses alone. Our work supports the claim that insertional mutagenesis is not required for in vitro reprogramming, and it provides a platform for studying the biology of iPS cells lacking viral integrations. For example, it should now be possible to assess if iPS cells and ES cells are equivalent at the molecular and functional levels. This comparison has not been possible so far because viral transgenes are expressed at low levels in iPS cells and their progeny, which may affect their molecular signatures, as well as their differentiation behavior and developmental potential. If human iPS cells can be generated without genome-integrating viruses, these cells may allow for the generation of safer patient-specific cells and thus could have important implications for cell therapy. Before translating these observations into a therapeutic setting, however, it will be important to assess if human iPS cells generated without viral integration are indeed as potent as human ES cells.

References and Notes

1. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).
2. M. Stadtfeld, K. Brennand, K. Hochedlinger, *Curr Biol* (2008).
3. J. Hanna *et al.*, *Cell* **133**, 250 (2008).
4. S. Eminli, J. Utikal, K. Arnold, R. Jaenisch, K. Hochedlinger, *Stem Cells*, published online 17 June 2008; in press.
5. N. Maherali *et al.*, *Cell Stem Cell* **3**, 340 (2008).
6. N. Maherali *et al.*, *Cell Stem Cell* **1**, 55 (2007).
7. W. E. Lowry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2883 (2008).
8. K. Okita, T. Ichisaka, S. Yamanaka, *Nature* **448**, 313 (2007).
9. K. Takahashi *et al.*, *Cell* **131**, 861 (2007).
10. M. Wernig *et al.*, *Nature* **448**, 318 (2007).
11. I. H. Park *et al.*, *Nature* **451**, 141 (2008).
12. J. Yu *et al.*, *Science* **318**, 1917 (2007).
13. M. Wernig *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5856 (2008).
14. J. Hanna *et al.*, *Science* **318**, 1920 (2007).
15. R. G. Hawley, *Mol. Ther.* **16**, 1354 (2008).
16. O. Kustikova *et al.*, *Science* **308**, 1171 (2005).
17. T. Aoi *et al.*, *Science* **321**, 699; published online 14 February 2008 (10.1126/science.1154884).
18. Materials and methods are available as supporting material on *Science* Online.
19. K. Hochedlinger, Y. Yamada, C. Beard, R. Jaenisch, *Cell* **121**, 465 (2005).
20. Q. Li, M. A. Kay, M. Finegold, L. D. Stratford-Perricaudet, S. L. Woo, *Hum. Gene Ther.* **4**, 403 (1993).
21. S. Yamada *et al.*, *Endocr. J.* **53**, 789 (2006).
22. A. Harui, S. Suzuki, S. Kochanek, K. Mitani, *J. Virol.* **73**, 6141 (1999).
23. N. Louis, C. Evelegh, F. L. Graham, *Virology* **233**, 423 (1997).
24. T. Brambrink *et al.*, *Cell Stem Cell* **2**, 151 (2008).
25. M. Stadtfeld, N. Maherali, D. T. Breault, K. Hochedlinger, *Cell Stem Cell* 10.1016/j.stem.2008.02.001 (2008).
26. D. Huangfu *et al.*, *Nat. Biotechnol.* **26**, 795 (2008).
27. T. S. Mikkelsen *et al.*, *Nature* **454**, 49 (2008).
28. Y. Shi *et al.*, *Cell Stem Cell* **2**, 525 (2008).
29. S. Gupta, *Semin. Cancer Biol.* **10**, 161 (2000).

30. We thank H. Hock for critical comments on the manuscript and L. Prickett and K. Folz-Donahue for expert help with flow cytometry. We are especially grateful to J. Lee of the Harvard Gene Therapy Initiative for help with the generation of adenoviral vectors. M.S. was supported by a Schering postdoctoral fellowship, J.U. was supported by a Mildred Scheel postdoctoral fellowship and M.N. and G.W. received support from the Juvenile Diabetes Research Foundation. Support to K.H. was from the NIH Director's Innovator Award, the Harvard Stem Cell Institute, the Kimmel Foundation and the V Foundation. The authors are filing a patent based on the results reported in this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1162494/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S3
References

30 June 2008; accepted 17 September 2008
Published online 25 September 2008;
10.1126/science.1162494
Include this information when citing this paper.

Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors

Keisuke Okita,[1] Masato Nakagawa,[1,2] Hong Hyenjong,[2] Tomoko Ichisaka,[1,3] Shinya Yamanaka[1,2,3,4]*

Induced pluripotent stem (iPS) cells have been generated from mouse and human somatic cells by introducing Oct3/4 and Sox2 with either Klf4 and c-Myc or Nanog and Lin28 using retroviruses or lentiviruses. Patient-specific iPS cells could be useful in drug discovery and regenerative medicine. However, viral integration into the host genome increases the risk of tumorigenicity. Here, we report the generation of mouse iPS cells without viral vectors. Repeated transfection of two expression plasmids, one containing the complementary DNAs (cDNAs) of Oct3/4, Sox2, and Klf4 and the other containing the c-Myc cDNA, into mouse embryonic fibroblasts resulted in iPS cells without evidence of plasmid integration, which produced teratomas when transplanted into mice and contributed to adult chimeras. The production of virus-free iPS cells, albeit from embryonic fibroblasts, addresses a critical safety concern for potential use of iPS cells in regenerative medicine.

• PS cells were first generated from mouse fibroblasts by retroviral-mediated introduction of four factors, Oct3/4, Sox2, Klf4, and c-Myc (*1*). Human fibroblasts can also be reprogrammed by the same four factors (*2*–*4*) or by Oct3/4, Sox2, Nanog, and Lin28 (*5*). Mouse and human iPS cells are similar to embryonic stem (ES) cells in morphology, gene expression, epigenetic status, and in vitro differentiation. Furthermore, mouse iPS cells give rise to adult chimeras and show competence for germline transmission (*6*–*8*). However, chimeras and progeny mice derived from iPS cells frequently develop tumors, which in some cases may be due to reactivation of the c-Myc oncogene (*7*). It is possible to generate iPS cells without retroviral insertion of c-Myc (*9*, *10*), albeit at a lower efficiency. Nevertheless, retroviral integration of the other transcription factors may activate or inactivate host genes, resulting in tumorigenicity, as was the case in some patients who underwent gene therapy (*11*). In order to apply the technology to cell transplantation therapy, it is crucial to generate iPS cells with use of nonintegration methods (*12*).

To generate mouse iPS cells without retroviruses, we used an adenovirus-mediated gene delivery system. As an initial step, we generated iPS cells with one or two factors using adenoviruses and using retroviruses for the remaining factors. We used mice in which green fluorescence protein (GFP) and the puromycin-

[1]Center for iPS Cell Research and Application (CiRA), Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 606-8507, Japan. [2]Department of Stem Cell Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan. [3]Core Research for Evolutional Science and Technology and Yamanaka iPS Cell Project, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. [4]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA.

*To whom correspondence should be addressed. E-mail: yamanaka@frontier.kyoto-u.ac.jp

resistant gene are driven by the *Nanog* enhancer and promoter (*7*). With the Nanog reporter, iPS cells can be selected with puromycin and detected as GFP-positive colonies. We transduced mouse primary hepatocytes from the Nanog reporter mice with combinations of retroviruses and adenoviruses. We chose hepatocytes because iPS cells derived from hepatocytes have fewer retroviral integration sites than do iPS cells derived from fibroblasts (*13*). Because transgene expression should be maintained for up to 12 days during iPS cell generation (*14*, *15*), we repeatedly delivered adenoviruses. We observed GFP-positive colonies when Sox2 or Klf4 was introduced with adenovirus and the remaining two factors–Oct3/4 and Klf4 or Oct3/4 and Sox2, respectively–were introduced with retroviruses (Fig. 1A). We confirmed that these iPS cells did not show integration of adenoviral transgenes (Fig. 1B). They expressed markers of ES cells, including Nanog, Rex1, and ECAT1, in quantities similar to those in ES cells (Fig. 1C). They formed teratomas containing derivatives of all three germ layers when transplanted subcutaneously into nude mice (Fig. 1D). No GFP-positive colonies emerged when Oct3/4 was introduced with adenoviruses and Klf4 and Sox2 with retroviruses (Fig. 1A). Furthermore, we did not obtain GFP-positive colonies upon introduction of two factors by adenoviruses.

We were unable to generate iPS cells by introducing the four factors with separate adenoviral vectors. This might be due to the inability to introduce multiple factors into the same cells at sufficient concentrations. Hence, we placed the cDNAs encoding Oct3/4, Sox2, and Klf4 into a single expression vector. To this end, we used the foot-and-mouth disease virus 2A self-cleaving peptide (*16*, *17*), which enables efficient polycistronic expression in ES cells. We first placed the three cDNAs in all possible orders into pMXs retroviral vectors (*18*) (fig. S1A). We then transduced Nanog reporter mouse embryonic fibroblasts (MEFs) with these retroviruses to induce iPS cells. We observed the highest efficiency of GFP-positive colony formation when the factors were in order as Oct3/4, Klf4, and then Sox2 (fig. S1, B and C).

Next, we placed the three factors in this same order into a plasmid vector containing the CAG constitutively active promoter (*19*) (pCX-OKS-2A, Fig. 2A). In addition, we constructed another plasmid to express c-Myc (pCX-cMyc, Fig. 2A). In the initial attempt (experiment number 432), we transfected pCX-OKS-2A on days 1 and 3 and pCX-cMyc on days 2 and 4 (Fig. 2B). We obtained GFP-positive colonies that were morphologically indistinguishable from mouse ES cells (Fig. 2C). These virus-free iPS cells expressed ES cell marker genes at levels comparable to those in ES cells (Fig. 2D) and gave rise to adult chimeric mice (Fig. 2E). These data showed that mouse iPS cells can be generated without retroviruses or lentiviruses. Polymerase chain reaction (PCR) analyses, however, detected plasmid incorporation into the host genome (Fig. 2F).

We then modified the transfection protocol in order to avoid plasmid integration. We transfected pCX-OKS-2A and pCX-cMyc together on days 1, 3, 5, and 7 (experiment number 440, Fig. 2B). We obtained multiple GFP-positive colonies, which gave rise to cells morphologically indistinguishable from ES cells (Fig. 2C). These cells expressed markers of ES cells at comparable levels (Fig. 2D). To test for genomic integration of plasmid DNA, we designed 16 sets of PCR

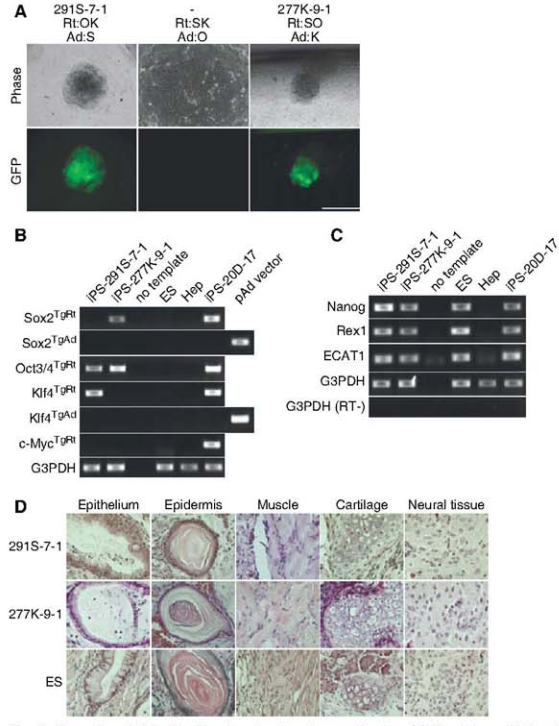


Fig. 1. Generation of iPS cells with adenovirus/retrovirus combination. **(A)** Morphology of iPS cells established by adenovirus/retrovirus combination. The iPS cell clone 291S-7-1 was generated with retroviral (Rt) introduction of Oct3/4 (O) and Klf4 (K) and adenoviral (Ad) introduction of Sox2 (S). The iPS cell clone 277K-9-1 was generated with retroviral introduction of Oct3/4 and Sox2 and adenoviral introduction of Klf4. Scale bar indicates 500 μm. **(B)** Integration of retroviral transgenes (TgRt) and adenoviral transgenes (TgAd) of the four factors was examined by genomic PCR analyses. As a control of TgRt, iPS cells generated with retroviruses (clone 20D-17) were used. As a control of TgAd, the plasmid pAd/CMV/V5-DEST vectors containing Sox2 or Klf4 were used. ES, RF8 mouse ES cells; Hep, mouse primary hepatocytes. As a loading control, the housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (G3PDH) was used. **(C)** Reverse transcription PCR (RT-PCR) analyses of ES cell marker genes and transgenes in RF8 mouse ES cells and iPS cells (clones 291S-7-1, 277K-9-1, and 20D-17). As a loading control, G3PDH was used. As a negative control, G3PDH was amplified without the reverse transcriptase (RT–). **(D)** Teratoma formation. RF8 ES cells or iPS cells (clones 291S-7-1 and 277K-9-1) were subcutaneously transplanted into nude mice. After 4 weeks, tumors were sectioned and stained with hematoxylin and eosin staining. Shown are gutlike epithelial tissues (left), epidermal tissues, striated muscles, cartilage, and neural tissues (right).

primers to amplify various parts of the plasmids (Fig. 2A). In 9 out of 11 GFP-positive clones obtained by the modified protocol, no amplification of plasmid DNA was observed (Fig. 2F). In addition, Southern blot analyses did not detect integration of transgenes in these clones (fig. S2). Although we cannot formally exclude the presence of small plasmid fragments, these data show that the iPS cells are most likely free from plasmid integration into the host genome.

To exclude the possibility that virus-free iPS cells were derived from contamination of Nanog reporter ES cells that we have in our laboratory, we performed simple sequence length polymorphisms (SSLP) analyses. In experiment number

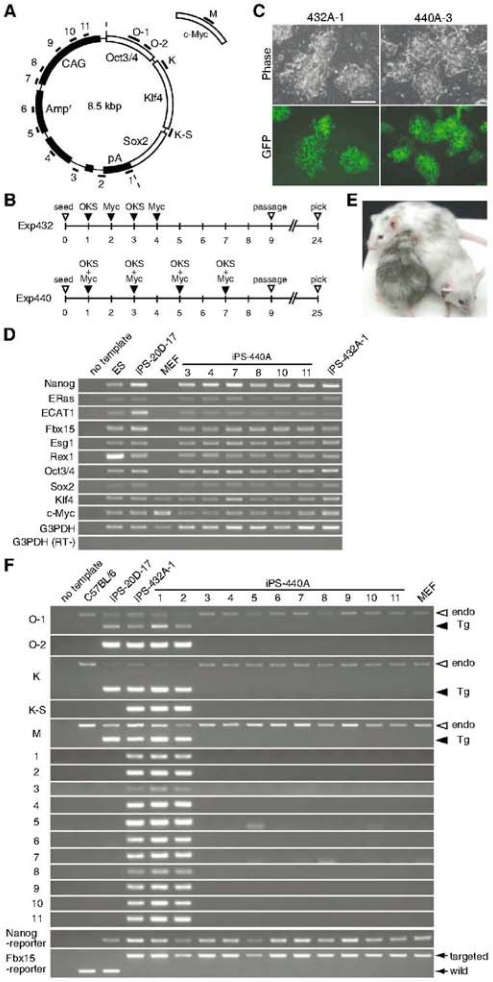


Fig. 2. Generation of virus-free iPS cells. **(A)** Expression plasmids for iPS cell generation. The three cDNAs encoding Oct3/4, Klf4, and Sox2 were connected in this order with the 2A peptide and inserted into the pCX plasmid (pCX-OKS-2A). In addition, the c-Myc cDNA was inserted into pCX (pCX-cMyc). Thick lines (O-1, O-2, K, K-S, 1 to 11, and M) indicate amplified regions used in (F) to detect plasmid integration into genome. The locations of the CAG promoter, the ampicillin-resistant gene (Amp^r), and the polyadenylation signal (pA) are also shown. **(B)** Time schedules for induction of iPS cells with plasmids. Open arrowheads indicate the timing of cell seed, passage, and colony pickup. Solid arrowheads indicate the timing of transfection. **(C)** Colonies of virus-free iPS cells. Scale bar, 200 μm. **(D)** Gene expression. Total RNAs isolated from ES cells, retrovirus-induced iPS cells (clone 20D-17), plasmid-induced iPS cells (clones 440A-3, -4, -7, -8, -10, and -11 and clone 432A-1), and MEFs were analyzed with RT-PCR. **(E)** Chimeric mice derived from the clone 432A-1. **(F)** Detection of plasmid integration by PCR. Genomic DNA from a C57BL/6 mouse, retrovirus-induced iPS cells (clone 20D-17), plasmid-induced iPS cells (clone 432A-1 and clones 440A-1 to -11), and MEFs were amplified by PCR with primers indicated in (A). In PCR for O-1, K, and M, bands derived from the endogenous (endo) genes are shown with open arrowheads, whereas those from integrated plasmids (Tg) shown with solid arrowheads. For the Fbx15 reporter, the lower bands indicate the wild-type allele, whereas the upper bands indicate the knock-in allele. Minor PCR bands seen in some virus-free clones are smaller than expected and are most likely derived from primer dimers.

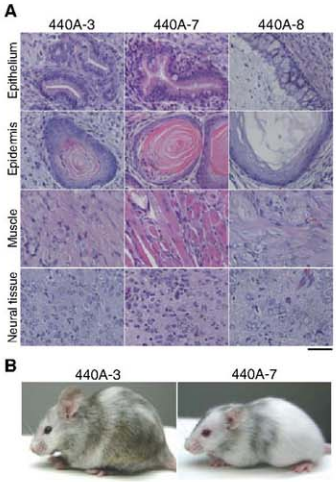


Fig. 3. Pluripotency of virus-free iPS cells without evidence of transgene integration. **(A)** Teratoma formation. Virus-free iPS cells (clones 440A-3, -7, and -8) were subcutaneously transplanted into nude mice. After 4 weeks, tumors were sectioned and stained with hematoxylin and eosin staining. Shown are guttlike epithelial tissues (upper), epidermal tissues, striated muscles, and neural tissues (bottom). Scale bar, 50 μm. **(B)** Chimeric mice derived from virus-free iPS cells (clones 440A-3 and -7). **(C)** Detection of plasmid integration by PCR. Genomic DNA was extracted from an ICR mouse, iPS cells (clone 432A-1), and chimeric mice derived from plasmid-induced iPS cells (clone 432A-1 and clones 440A-3, -7, -8, and -11) and was analyzed with PCR to amplify fragments O-1, K, and M indicated in Fig. 2A. Bands derived from the endogenous genes are shown with open arrowheads, whereas those from integrated plasmids were with solid arrowheads. The presence of the Nanog reporter and the Fbx15 reporter was also detected by PCR.

440, we used MEFs derived from five mouse embryos. SSLP analyses were able to distinguish these five and identified the origin of each virus-free iPS cells. The analyses also confirmed that virus-free iPS cells are different from ES cells, the latter of which were derived from the mouse strain 129S4 (fig. S3).

We performed 10 independent experiments with this transfection protocol (table S1). In 7 out of the 10 experiments, we obtained GFP-positive colonies. In these experiments, 1 to 29 GFP-positive colonies emerged from 1×10^6 transfected cells. When we used retroviruses, we routinely obtained >100 GFP-positive colonies with this number of transfected cells when using three factors and ~1000 GFP-positive colonies with the four factors. In addition, because we re-plated transfected MEFs at day 9, some virus-free iPS cell clones may derive from common progenitor cells. Thus, the efficiency of iPS cell induction with the plasmid transfection protocol is substantially lower than that with the retroviral method. Nevertheless, we obtained iPS cell clones without evidence of plasmid integration in 6 out of 10 experiments (figs. S2 to S5), demonstrating reproducibility of the protocol.

To confirm pluripotency of these iPS cells, we transplanted them subcutaneously into nude mice. All clones tested (440A-3, -4, -7, -8, -10, and -11; 492B-4 and -9; 497B-30; and 497D-2) gave rise to tumors containing a wide variety of cell types, including cells derived from all three germ layers (Fig. 3A and fig. S5). We also injected the iPS cells into blastocysts from the mouse strain ICR. From all clones injected (440A-3, -4, -6, -7, -8, -9, -10, and -11; 492B-4 and 9), we obtained adult chimeras, as judged from coat color (Fig. 3B). In these chimera mice, we did not detect integration of the transgenes by PCR (Fig. 3C). Furthermore, PCR analyses detected both the Nanog reporter and the Fbx15 reporter (*20*) in chimeras (Fig. 3C). Because we generated these virus-free iPS cells from the double reporter mice in the experiment number 440 and because we do not have double reporter ES cells in our laboratory, these data confirmed that these chimeras were derived from the virus-free iPS cells but not from contaminated ES cells. These results confirm pluripotency of iPS cells generated by plasmid transfection methods.

We previously reported that we did not find common retroviral integration sites in iPS cell derived from mouse liver and stomach (*13*). The current study shows dispensability of retroviral integration in iPS cell generation. The efficiency of iPS cell generation, however, is substantially lower without retroviruses. This may suggest that retroviral integration facilitates iPS cell generation. Alternatively, the lower efficiency may be attributable to lower transgene expression levels observed with plasmid transfection than those with retroviruses (fig. S6). Further studies are required to increase the efficiency of virus-free iPS cells. In addition, whether virus-free iPS cells are germline-competent and whether they can be generated from adult somatic cells remain to be determined. Nevertheless, our study is an important step toward studying patient-specific cells and associated disease as well as future application of iPS cell technology in regenerative medicine and other clinical usages.

References and Notes

1. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).
2. W. E. Lowry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2883 (2008).
3. J. H. Park *et al.*, *Nature* **451**, 141 (2008).
4. K. Takahashi *et al.*, *Cell* **131**, 861 (2007).
5. J. Yu *et al.*, *Science* **318**, 1917 (2007); published online 19 November 2007 (10.1126/science.1151526).
6. N. Maherali *et al.*, *Cell Stem Cell* **1**, 55 (2007).
7. K. Okita, T. Ichisaka, S. Yamanaka, *Nature* **448**, 313 (2007).
8. M. Wernig *et al.*, *Nature* **448**, 318 (2007).
9. M. Nakagawa *et al.*, *Nat. Biotechnol.* **26**, 101 (2008).
10. M. Wernig, A. Meissner, J. P. Cassady, R. Jaenisch, *Cell Stem Cell* **2**, 10 (2008).
11. A. W. Nienhuis, C. E. Dunbar, B. P. Sorrentino, *Mol. Ther.* **13**, 1031 (2006).
12. S. Yamanaka, *Cell Stem Cell* **1**, 39 (2007).
13. T. Aoi *et al.*, *Science* **321**, 699 (2008); published online 12 February 2008 (10.1126/science.1154884).
14. T. Brambrink *et al.*, *Cell Stem Cell* **2**, 151 (2008).
15. M. Stadtfeld, N. Maherali, D. T. Breault, K. Hochedlinger, *Cell Stem Cell* **2**, 230 (2008).
16. E. C. Hsiao *et al.*, *PLoS One* **3**, e2532 (2008).
17. K. Hasegawa, A. B. Cowan, N. Nakatsuji, H. Suemori, *Stem Cells* **25**, 1707 (2007).
18. S. Morita, T. Kojima, T. Kitamura, *Gene Ther.* **7**, 1063 (2000).
19. H. Niwa, K. Yamamura, J. Miyazaki, *Gene* **108**, 193 (1991).
20. Y. Tokuzawa *et al.*, *Mol. Cell. Biol.* **23**, 2699 (2003).

21. We thank K. Takahashi, K. Yae, M. Koyanagi, and K. Tanabe for scientific discussion; K. Okuda, M. Narita, A. Okada, H. Miyachi, S. Kitano, and N. Takizawa for technical assistance; R. Kato, R. Iyama, E. Nishikawa, and Y. Shimazu for administrative assistance; and D. Srivastava for critical reading of the manuscript. We also thank J. Miyazaki for the CAG promoter.

T. Kitamura for Plat-E cells and pMXs retroviral vectors, and R. Farese for RF8 ES cells. This study was supported in part by a grant from the Program for Promotion of Fundamental Studies in Health Sciences of National Institute of Biomedical Innovation (NIBIO), a grant from the Leading Project of Ministry of Education, Culture, Sports, Science, and Technology (MEXT), a grant from Uehara Memorial Foundation, and grants-in-aid for scientific research of Japan Society for the Promotion of Science (JSPS) and MEXT (to S.Y.). K.O. was a JSPS research fellow. H.H. is supported by a Japanese government (MEXT) scholarship. The authors are filing a patent based on the results reported in this paper.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1164270/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2

6 August 2008; accepted 25 September 2008
Published online 9 October 2008;
10.1126/science.1164270
Include this information when citing this paper.

Insights into Translational Termination from the Structure of RF2 Bound to the Ribosome

Albert Weixlbaumer,* Hong Jin,* Cajetan Neubauer, Rebecca M. Voorhees, Sabine Petry,† Ann C. Kelley, Venki Ramakrishnan‡

The termination of protein synthesis occurs through the specific recognition of a stop codon in the A site of the ribosome by a release factor (RF), which then catalyzes the hydrolysis of the nascent protein chain from the P-site transfer RNA. Here we present, at a resolution of 3.5 angstroms, the crystal structure of RF2 in complex with its cognate UGA stop codon in the 70*S* ribosome. The structure provides insight into how RF2 specifically recognizes the stop codon; it also suggests a model for the role of a universally conserved GGQ motif in the catalysis of peptide release.

In nearly all species, three stop codons, UGA, UAG, and UAA, signal the end of the coding sequence in mRNA. These stop codons are decoded by a protein factor termed a class I release factor (RF) (*1*, *2*). In bacteria, there are two such factors with overlapping specificity: RF1 recognizes UAG, RF2 recognizes UGA, and both recognize UAA. In eukaryotes, a single RF, eRF1, recognizes all three stop codons. The mechanism by which RFs specifically decode stop codons and catalyze peptidyl-tRNA hydrolysis is a fundamental problem in understanding translation.

Elements of RFs involved in catalysis and stop-codon recognition have been proposed, using sequence analysis combined with biochemistry and genetics. A universally conserved tripeptide sequence, GGQ (*3*), has been implicated in the hydrolysis of the peptide chain from tRNA (*4*). Exchanging a tripeptide motif between RF1 and RF2 [P(A/V)T in RF1; SPF in RF2] switches their respective specificities for UAG and UGA (*5*). Hydroxyl-radical probing suggested that the SPF and GGQ motifs were close to the decoding center and peptidyl transferase center (PTC), respectively (*6*).

The structure of the eukaryotic eRF1 suggested that the distance between its codon-recognition and GGQ motifs was compatible with the approximately 75 Å distance between the decoding center and the PTC (*7*). However, eRF1 has no sequence or structural homology to bacterial RFs. The crystal structure of a bacterial RF2 (*8*) showed that the distance between the SPF and GGQ motifs was about 23 Å and thus incompatible with their simultaneous involvement in decoding and peptide release. This anomaly was resolved when low-resolution structures (*9*–*11*) showed that the conformation of RF2 when bound to the ribosome was different, so that the GGQ and SPF/PVT motifs were localized to the PTC and decoding center, respectively. However, the resolution of these

Medical Research Council (MRC) Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

*These authors contributed equally to this work.
†Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158-2517, USA.
‡To whom correspondence should be addressed. E-mail: ramak@mrc-lmb.cam.ac.uk

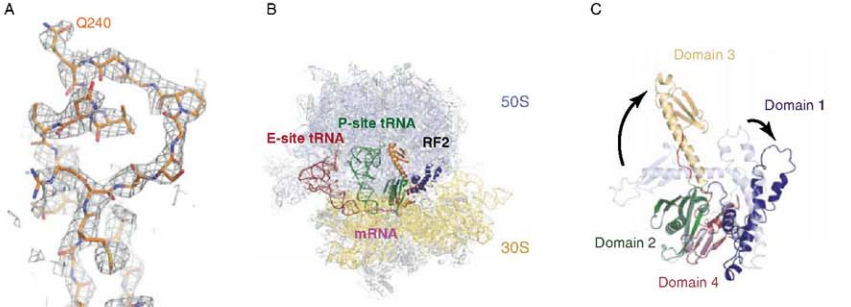


Fig. 1. The structure of RF2 in the ribosome. **(A)** Unbiased difference Fourier maps showing the density of RF2 in the peptidyl transferase center of the ribosome. **(B)** Overview of the structure showing the 50*S* subunit (light blue), the 30*S* subunit (yellow) with E-site tRNA (red), P-site tRNA (green), and RF2 colored by domains as in (C). The mRNA is shown in magenta. **(C)** Conformational differences between the isolated crystal structure of RF2 (*15*) (shown in light blue), and the ribosome-bound form (colored by domains as labeled) are indicated. Loop regions connecting domain 3 with domains 2 and 4, that undergo substantial conformational changes, are highlighted in red.

structures was too low to reveal detailed interactions of RFs with the ribosome.

Two years ago, our laboratory identified a crystal form of the *Thermus thermophilus* ribosome that diffracts to high resolution in the presence of both A- and P-site ligands (*12*), thus providing a way out of this impasse. Using this crystal form, we report a structure, refined to a resolution of 3.45 Å, of RF2 bound to a 70*S* ribosome containing the RF2-specific UGA stop codon (whose nucleotides are named U1, G2, and A3) in the A site (*13*). The structure shows details of stop-codon recognition by RF2 as well as interactions of RF2 with the PTC and other regions of the ribosome, including helix 69 (H69) and the L11 region. This crystal form was also used in a recent structure of RF1 bound to the ribosome containing a UAA codon (*14*).

Initial refinement of the data by using a model of the empty ribosome showed clear difference density for the bound ligands (Fig. 1A). A complete model for the ribosome in complex with P- and E-site $tRNA^{Phe}$, mRNA, and RF2 was thus built and refined (*13*) (Fig. 1B). Compared with the crystal structure of isolated *T. thermophilus* RF2 (*15*), the factor in complex with the ribosome underwent major conformational changes. As expected from earlier low-resolution studies (*9*–*11*), domain 3 of RF2 peeled away from domains 2 and 4 (*15*) (Fig. 1C), placing the GGQ motif in the PTC, whereas domains 2 and 4 interact with the decoding center. Domain 1 shifted slightly and interacts with the L11 region of the 50*S* subunit (see fig. S1 for details). These global conformational changes were accompanied by rearrangements of specific regions (Fig. 1C).

At the decoding center, RF2 induces changes that are markedly different from those induced by paromomycin or tRNA binding to sense codons (*16*, *17*) (Fig. 2A). In standard decoding, three conserved nucleotides in the 30*S* ribosomal subunit, A1492, A1493, and G530 (*Escherichia coli* numbering is used throughout), change conformation upon tRNA binding and closely interact with the minor groove of the codon-anticodon helix. These conformational changes would clash with RF2, explaining why paromomycin promotes tRNA binding but abolishes RF binding (*18*). In its new orientation, A1493 stacks

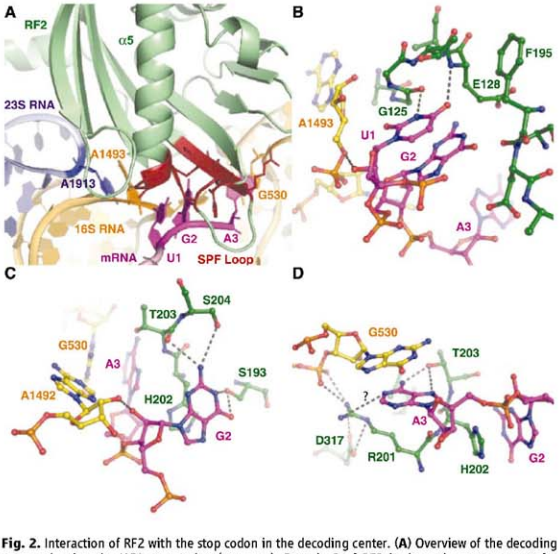


Fig. 2. Interaction of RF2 with the stop codon in the decoding center. **(A)** Overview of the decoding center, showing the UGA stop codon (magenta). Domain 2 of RF2 is shown in green, except for parts interacting with the codon, which are shown in red. Key bases from 16*S* RNA and A1913 of 23*S* RNA are also shown. **(B** to **D)** Details of interactions at the first (B), second (C), and third (D) positions of the stop codon with elements of RF2 and the decoding center.

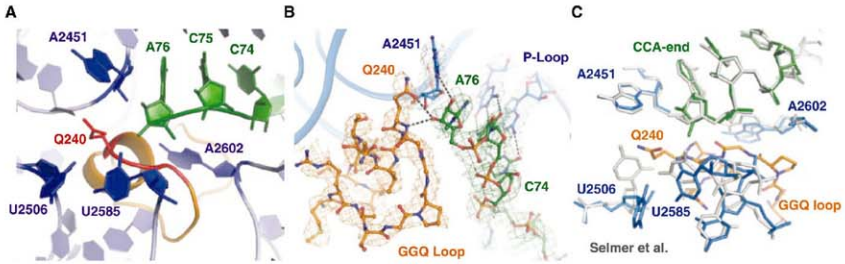


Fig. 3. Interaction of RF2 with the peptidyl transferase center. **(A)** Overview of the peptidyl transferase center showing the terminal CCA of P-site tRNA (green), the conserved GGQ motif of RF2 (red), and key bases of 23*S* RNA (blue). **(B)** Interaction of the GGQ loop of RF2 (orange) with the terminal ribose of P-site tRNA (green), showing σ_A weighted $3mF_{obs} - 2DF_{calc}$ maps, where m is the figure of merit, D is the σ_A weight, and F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. Potential hydrogen bonds between the conserved Q240 and the ribose of P-site tRNA are shown. **(C)** Changes in the peptidyl transferase center upon RF2 binding compared with the 70*S* structure with an empty A site (gray) (*12*). The structure shows that RF2 would clash with elements of 23*S* RNA and induces conformational changes, in particular in U2506 and U2585.

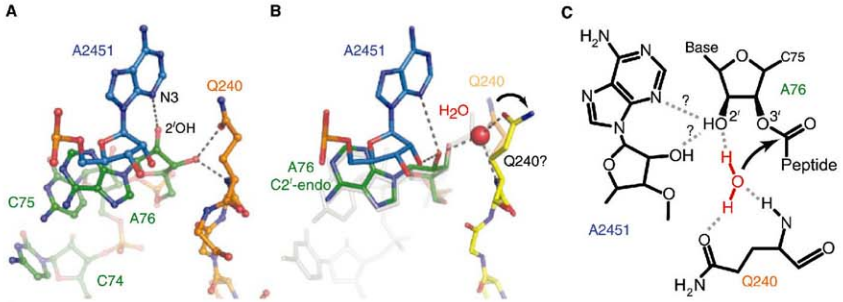


Fig. 4. A model of the substrate complex that suggests the basis for catalysis. **(A)** The interaction of Q240 of RF2 and A2451 of 23*S* RNA in the current structure. **(B)** A proposed structure showing how a minor change in the orientation of Q240 would allow it to coordinate a water molecule (transparent orange versus yellow). In gray is a transition-state analog superposed on this structure [1VQ7, taken from (*24*)] that was used to place the putative water that would take part in a nucleophilic attack on the ester bond. **(C)** A schematic representation of how a network of interactions among the conserved Q240, the coordinated water molecule that is the attacking nucleophile, the ribose of A76 of P-site tRNA (in C2′-endo conformation), and A2451 would act to facilitate catalysis.

on A1913 of 23*S* RNA, which probably plays a role in signal transduction from the decoding center to the PTC.

The structure sheds light on the basis of stop-codon recognition by RF2. At the first position, two conserved glycines at the tip of helix α5 in domain 2 make important backbone contacts with U1 of the stop codon (Fig. 2B). A purine at this position would result in a steric clash with the backbone of both G125 and G126. The N3 of U1 can make a hydrogen bond with the carbonyl group of G125, and the O4 is within hydrogen-bonding distance of the backbone amide of E128. These interactions would not occur with a C. Together with the recent structure of RF1 bound to the ribosome (*14*), these observations explain why U is required at the first position of all stop codons.

Only the first amino acid in the SPF motif directly interacts with the second base of the stop codon, via hydrogen bonds between S193 and the O6 and N1 of G2 (Fig. 2C). Similar interactions with the hydroxyl group of S193 would be possible with the N6 and N1 of A, which suggests why an A would also be recognized at this position. The imidazole ring of H202 makes a stacking interaction with the second base. Though a pyrimidine would stack with H202 in a similar fashion, complete desolvation of the Watson-Crick edge would be required without the formation of compensatory hydrogen bonds. Additional conserved elements are important for stop-codon recognition: both the hydroxyl of S204 and the carbonyl backbone of T203 can make hydrogen bonds with the N2 of G2.

Discrimination at the third position appears to be more complex. Although both RFs recognize an A at this position, RF2 miscodes a UGG codon with an error rate of 1 out of 2400, which is almost an order of magnitude worse than it miscodes UGU or UGC (*19*). The third base A3 does not stack on G2 but instead on G530 of 16S RNA (Fig. 2D). This stacking interaction would be far less favorable for a pyrimidine, which explains the discrimination against U or C at this position. The hydroxyl group of T203 can donate a proton to the N7 of A3 and accept one from its N6. With a G, similar hydrogen bonds could not be formed simultaneously to both its O6 and N7. However, T203 is also present in RF1, which can accept a G at the third position, suggesting that other interactions monitor this position. The terminal amino group of R201, which is conserved in RF2 but absent in RF1, is within hydrogen-bonding distance of the phosphate of U531 and D317 in domain 4 of RF2. These interactions could position the arginine to act as a hydrogen-bond donor for the N1 of A3, which would not be possible with a G. However, the arginine is poorly ordered in the structure.

The structure thus provides a rationale for the specific recognition of UAA and UGA by RF2 and, in conjunction with the recent structure of RF1, clarifies many of the elements involved in stop-codon recognition. However, it is clear that the tripeptide motif identified genetically (*5*) is only partially responsible for recognition. Other residues, some of which have been discussed, are probably involved in conferring specificity. Moreover, mutations in RF2 distant from the codon give rise to bacterial RFs with altered specificity (*20*). This suggests that recognition and miscoding by RFs are more complex and may involve the kind of tradeoff between binding energy and induced conformational changes that has been observed for tRNA recognition (*21*). The observation that the accuracy of termination arises from both a binding component and a catalytic rate component (*19*) is also fully consistent with the idea that codon recognition by RFs is followed by induced structural rearrangements that lead to catalysis.

The GGQ motif essential for catalysis is positioned in the heart of the PTC (Fig. 3A). The loop within domain 3 that contains this conserved tripeptide shifts considerably when bound to the ribosome as compared with the isolated crystal structure (*15*). In the modeled loop, both glycines adopt a backbone conformation disallowed for any other amino acid, which explains their universal conservation and the drastic reduction in in vitro RF activity upon their mutation (*22*, *23*). The glutamine side chain Q240 was modeled and refined in an orientation that is consistent with the initial unbiased difference Fourier maps (Fig. 1A), which places its side-chain carbonyl within hydrogen-bonding distance of the ribose of A76 of P-site tRNA (Fig. 3B). After refinement, difference Fourier maps also suggest the existence of a second conformation similar to that in the recent structure of RF1 bound to the ribosome, where it is pointed away from the ribose (*14*). The resolution is not high enough to distinguish between the C2′- and C3′-endo conformations of the ribose. In Fig. 3B, the ribose was modeled in a C3′-endo conformation, which in addition to a bond between the glutamine oxygen to the 3′-OH of A76 also allows a bond between the backbone amide of Q240 and the 3′-OH of A76. This backbone interaction was observed with RF1 (*14*), where it has been suggested to be important for product stabilization. The alternative C2′-endo conformation seen in other structures (*12*, *24*) would still allow the glutamine side-chain oxygen to make hydrogen bond with the 2′-OH of A76 but would preclude formation of a hydrogen bond with the main-chain amide.

RF2 also induces changes in the PTC as compared with the structure of the ribosome with an empty 50S A site (*12*). The GGQ loop would clash with U2506 and U2585 if they remained in

the same conformation as when the A site is unoccupied. Thus, both residues have moved as a result of RF2 binding (Fig. 3C). As suggested previously (*24*), the movement of U2585 away from the ester bond of peptidyl tRNA opens the bond to nucleophilic attack by water, whereas in the uninduced state it would be protected by U2585 in the absence of an A-site ligand. The changes induced by the binding of the GGQ loop that expose the ester bond to nucleophilic attack must be a major component of the catalytic mechanism of RFs and, as suggested previously, show how similar changes induced by deacylated tRNA binding can catalyze peptide release (*24*).

The glutamine side chain was proposed to directly coordinate a water molecule during hydrolysis of the peptidyl–tRNA ester bond (*7*). Its mutation had a less drastic effect on activity than mutating either glycine (*22*, *23*); however, the rate enhancement due to the glutamine must nevertheless be important because it is universally conserved, and RFs containing a glutamine mutation fail to complement defective RFs in vivo (*25*).

The structure, which is of the state after peptide release (Fig. 4A), makes it possible to propose a model for catalysis that rationalizes the available biochemical data by using substrate and transition-state analog structures. The superposition of a transition-state analog of peptidyl transfer from a 50*S* subunit complex (*24*) shows that only a small change in the dihedral angles of the conserved Q240 would be sufficient to accommodate a water molecule positioned for nucleophilic attack on the peptidyl–tRNA ester bond (Fig. 4B). As in models presented earlier (*7*, *26*), the side-chain oxygen of Q240 could be involved in hydrogen bonding with a water molecule (Fig. 4C). The glutamine is normally methylated in vivo, and the methylation is known to stimulate termination (*27*). Presumably, the methylation would help direct the amine away from the ribose, thus positioning the oxygen for coordination with the water molecule. By acting as a hydrogen bond donor to N3 of A2451 (or its 2′-OH), the 2′-OH of A76 (modeled in the C2′-endo conformation, as seen in previous studies) is positioned to accept the second hydrogen of the attacking water, positioning the water oxygen for inline attack on the ester. Our proposed model also explains the minimal effect of A2451 mutants on peptide release (*28*) because all four bases contain a hydrogen-bond acceptor in the position of the N3 of A2451 and have a 2′-OH. Furthermore, the involvement of the 2′-OH of A76 agrees with data showing a substantial effect on the peptide hydrolysis of a 2′-deoxyadenosine mutant at this position (*29*). The specific coordination of a water by glutamine is consistent with studies showing that alternative nucleophiles are not affected by its mutation (*23*). A recent molecular-dynamics study also proposes that the glutamine oxygen directly coordinates a water molecule, and suggests that mutation to alanine allows a second water molecule to compensate for the glutamine oxygen (*26*). Finally, the mechanism and the involvement of the 2′-OH of the peptidyl tRNA have similarities with the peptidyl transferase reaction (*23*).

Mutations of A2602 were observed to reduce the rate of peptide release (*28*), but other experiments reported that an abasic 2602 had no effect on the reaction rate (*30*). The structure shows that a direct involvement of A2602 in catalysis can be excluded, but it is possible that the nucleotide is involved in stabilizing the conformation of the GGQ loop, as has also been proposed for RF1 (*14*). Even so, the structure cannot explain why a mutation to G would result in a 90-fold reduced reaction rate (*28*).

In the ribosome, the closed form of RF2 would clash with P-site tRNA and H69, which is consistent with the observation that *E. coli* RF1 is primarily in the open form in solution (*31*). Although *Thermus* RF2 is mainly in the closed form at 20°C, thermodynamic data suggested that it too exists predominantly in the open form at the physiological temperature of 75°C (*15*). Thus, it is unlikely that RF2 binds in the closed form to the ribosome and switches into the open form as a result of codon recognition. More likely, the open form is conformationally variable, and codon recognition imposes restraints on domain 3 that place the GGQ motif in the correct position at the PTC. As also noted for RF1 (*14*), local rearrangements of specific regions of RF2 are induced by ribosome binding (Fig. 1C).

Helix 69 of the 50*S* subunit is in a different orientation relative to the structure, with a partially ordered tRNA in the A site (fig. S2) (*12*). Upon stop-codon recognition, A1913 of this helix stacks onto A1493 (Fig. 2A and fig. S2), thus establishing a connection between the decoding center and the 50*S* subunit. Also, Q121 of RF2 can make hydrogen bonds with both the 2′-OH of C1914 and the phosphate backbone of U1915. This is consistent with previous biochemical data that shows that the pseudouridylation (*32*) or deletion (*33*) of H69 can affect translation termination. A similar change was also seen with RF1 (*14*), although the details of the A1913/A1493 stack appear different.

The high-resolution structure of RF2 bound to the ribosome with its UGA stop codon provides a structural basis for understanding its role in translational termination. Together with the recently published structure of the RF1-ribosome complex (*14*), these structures represent a major advance in understanding both stop-codon recognition and peptide release. Stop-codon recognition appears to have many components, and the remaining factor-codon combinations, along with mutagenesis studies, will help clarify additional elements of recognition. We have proposed a model for catalysis that is consistent with the structure and biochemical data. However, both this and the RF1 structure represent the state after hydrolysis and peptide release, with a deacylated tRNA in the P site. Structures of substrate analog complexes at a resolution at which ordered water molecules would be visible, along with more directed biochemical experiments, will therefore be essential for understanding catalysis by these factors.

References and Notes

1. M. R. Capecchi, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 1144 (1967).
2. E. M. Youngman, M. E. McDonald, R. Green, *Annu. Rev. Microbiol.* **62**, 353 (2008).
3. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
4. L. Y. Frolova *et al.*, *RNA* **5**, 1014 (1999).
5. K. Ito, M. Uno, Y. Nakamura, *Nature* **403**, 680 (2000).
6. K. S. Wilson, K. Ito, H. F. Noller, Y. Nakamura, *Nat. Struct. Biol.* **7**, 866 (2000).
7. H. Song *et al.*, *Cell* **100**, 311 (2000).
8. B. Vestergaard *et al.*, *Mol. Cell* **8**, 1375 (2001).
9. U. B. Rawat *et al.*, *Nature* **421**, 87 (2003).
10. B. P. Klaholz *et al.*, *Nature* **421**, 90 (2003).
11. S. Petry *et al.*, *Cell* **123**, 1255 (2005).
12. M. Selmer *et al.*, *Science* **313**, 1935 (2006).
13. Materials and methods are available as supporting material on *Science* Online.
14. M. Laurberg *et al.*, *Nature* **454**, 852 (2008).
15. G. Zoldak *et al.*, *Nucleic Acids Res.* **35**, 1343 (2007).
16. J. M. Ogle *et al.*, *Science* **292**, 897 (2001).
17. J. M. Ogle, F. V. Murphy, M. J. Tarry, V. Ramakrishnan, *Cell* **111**, 721 (2002).
18. E. M. Youngman, S. L. He, L. J. Nikstad, R. Green, *Mol. Cell* **28**, 533 (2007).
19. D. V. Freistroffer, M. Kwiatkowski, R. H. Buckingham, M. Ehrenberg, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2046 (2000).
20. K. Ito, M. Uno, Y. Nakamura, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8165 (1998).
21. J. M. Ogle, V. Ramakrishnan, *Annu. Rev. Biochem.* **74**, 129 (2005).
22. A. V. Zavialov, L. Mora, R. H. Buckingham, M. Ehrenberg, *Mol. Cell* **10**, 789 (2002).
23. J. J. Shaw, R. Green, *Mol. Cell* **28**, 458 (2007).
24. T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz, *Nature* **438**, 520 (2005).
25. L. Mora *et al.*, *Mol. Microbiol.* **47**, 267 (2003).
26. S. Trobro, J. Aqvist, *Mol. Cell* **27**, 758 (2007).
27. V. Dincbas-Renqvist *et al.*, *EMBO J.* **19**, 6900 (2000).
28. E. M. Youngman, J. L. Brunelle, A. B. Kochaniak, R. Green, *Cell* **117**, 589 (2004).
29. J. L. Brunelle, J. J. Shaw, E. M. Youngman, R. Green, *RNA* **14**, 1526 (2008).
30. M. Amort *et al.*, *Nucleic Acids Res.* **35**, 5130 (2007).
31. B. Vestergaard *et al.*, *Mol. Cell* **20**, 929 (2005).
32. M. Ejby, M. A. Sorensen, S. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19410 (2007).
33. I. K. Ali, L. Lancaster, J. Feinberg, S. Joseph, H. F. Noller, *Mol. Cell* **23**, 865 (2006).

34. We thank M. Schmeing for his useful comments and M. Fuchs and C. Schulze-Briese for their help and advice with data collection at the Swiss Light Source. This work was supported by MRC UK, a program grant from the Wellcome Trust, and awards from the Agouron Institute and Louis-Jeantet Foundation. C.N. is supported by a Boehringer-Ingelheim fellowship and R.M.V. is supported by a Gates-Cambridge fellowship. Coordinates for the structure have been deposited in the Protein Data Bank with accession codes 2j15 to 2j18. V.R. holds stock options in Rib-X Pharmaceuticals, a company that develops antibacterial drugs that target the ribosome.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/953/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References

19 August 2008; accepted 26 September 2008
10.1126/science.1164840

Fat Metabolism Links Germline Stem Cells and Longevity in *C. elegans*

Meng C. Wang, Eyleen J. O'Rourke, Gary Ruvkun*

Fat metabolism, reproduction, and aging are intertwined regulatory axes; however, the mechanism by which they are coupled remains poorly understood. We found that germline stem cells (GSCs) actively modulate lipid hydrolysis in *Caenorhabditis elegans*, which in turn regulates longevity. GSC arrest promotes systemic lipolysis via induction of a specific fat lipase. Subsequently, fat mobilization is promoted and life span is prolonged. Constitutive expression of this lipase in fat storage tissue generates lean and long-lived animals. This lipase is a key factor in the lipid hydrolysis and increased longevity that are induced by decreased insulin signaling. These results suggest a link between *C. elegans* fat metabolism and longevity.

A balance of fat storage and mobilization is a universal feature of animal physiology (*1*). Reproduction is an energy-intensive process, which is modulated by the availability of nutrients and in turn influences lipid metabolism (*2*). Reproductive ability declines with age, and many organisms undergo reproductive senescence (*3*). Obesity increases with age and is also associated with the transition to menopause in women (*4*). Genetic studies have suggested endocrine roles of adipose tissue and the reproductive system in regulation of life span (*5*–*8*). Thus, understanding the mechanisms by which fat metabolism is coupled to reproductive cues may reveal systemic regulation of fat metabolism and provide insights into the control of aging.

Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02114, USA.

*To whom correspondence should be addressed. E-mail: ruvkun@molbio.mgh.harvard.edu

In *C. elegans*, the energetic demands of progeny production are profound. The gonad undergoes many more mitoses than does somatic tissue, and the biomass of the oocytes produced is approximately equal to the biomass increase from egg to adult. Thus, in the absence of reproduction, a surfeit of available energy could lead to an increase in fat storage. To test this idea, we ablated the precursor cells of the germ line in *C. elegans* with the use of a laser microbeam. The vital dye Nile Red was used to visualize fat storage droplets in living animals (*9*). Opposite to the expected increase in fat storage, germ line–ablated animals stored 50% as much fat as untreated animals (Fig. 1, A to C). This finding suggested a regulatory mechanism coupling reproduction and fat metabolism.

Fat storage is also aberrant in the sterile mutants *glp-1(e2141ts)* and *glp-4(bn2ts)*, which are defective in germline proliferation (*10*, *11*). The *glp* mutants showed a 50% decrease in fat storage at the nonpermissive temperature relative to the wild type (*N2*) (Fig. 1, D to G). A similar decrease was observed by staining with a BODIPY-labeled fatty acid analog (fig. S1) (*12*) or Sudan Black, a fat-specific dye (fig. S2) (*13*). At the permissive temperature, the *glp* mutants reproduced normally and their fat storage was similar to that of the wild type (fig. S3).

Fat storage can be altered by changes in either energy input or expenditure. Food intake and retention in the gut are unchanged in *glp-1* (Fig. 1, H and J); the food absorption rate is also normal (Fig. 1I). Normal locomotion in *glp-1* suggests that less fat storage is not due to an increase in physical activity (Fig. 1K). Therefore, decreased fat storage in the germ line–defective mutants is unlikely to be the result of alterations in energy intake and/or physical activity, and more likely reveals an altered endocrine signaling axis.

Production of vitellogenin-rich oocytes is the most energy-intensive reproductive function. We used *fem-3* sterile mutants to examine whether gametogenesis influences fat storage. The gain-of-function allele *fem-3(q20ts)* produces only sperm, whereas the loss-of-function allele *fem-3(e2006ts)* produces only oocytes at the nonpermissive temperature (*14*, *15*). Neither *fem-3* mutant exhibited abnormal lipid accumulation (Fig. 2, A to D). This result excludes the possibility that gametogenesis regulates fat storage, and it also suggests that sterility per se does not cause a change in lipid accumulation.

To test whether germline proliferation regulates fat storage, we shifted *glp-1* mutants to the restrictive temperature at different developmental stages to arrest germline proliferation at distinct points. Adults that are generated from L2 (early) temperature shifts carry few mitotic germ cells, whereas adults from L4 (late) temperature shifts form the germ line with essentially wild type–

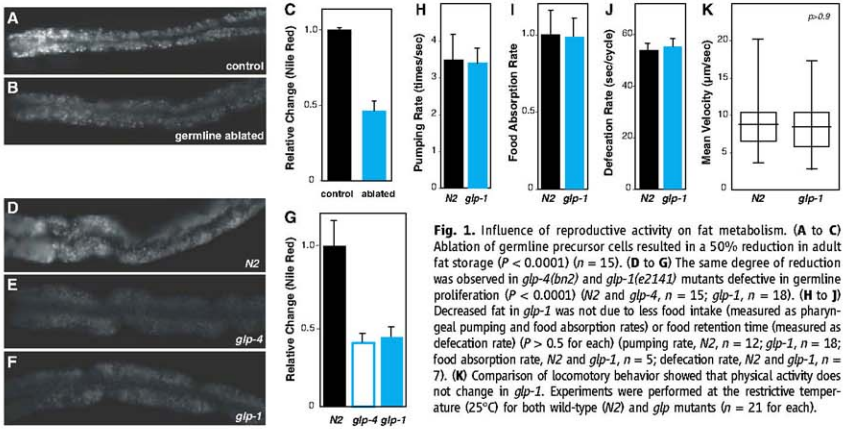


Fig. 1. Influence of reproductive activity on fat metabolism. (**A** to **C**) Ablation of germline precursor cells resulted in a 50% reduction in adult fat storage ($P < 0.0001$) ($n = 15$). (**D** to **G**) The same degree of reduction was observed in *glp-4(bn2)* and *glp-1(e2141)* mutants defective in germline proliferation ($P < 0.0001$) (*N2* and *glp-4*, $n = 15$; *glp-1*, $n = 18$). (**H** to **J**) Decreased fat in *glp-1* was not due to less food intake (measured as pharyngeal pumping and food absorption rates) or food retention time (measured as defecation rate) ($P > 0.5$ for each) (pumping rate, *N2*, $n = 12$; *glp-1*, $n = 18$; food absorption rate, *N2* and *glp-1*, $n = 5$; defecation rate, *N2* and *glp-1*, $n = 7$). (**K**) Comparison of locomotory behavior showed that physical activity does not change in *glp-1*. Experiments were performed at the restrictive temperature (25°C) for both wild-type (*N2*) and *glp* mutants ($n = 21$ for each).

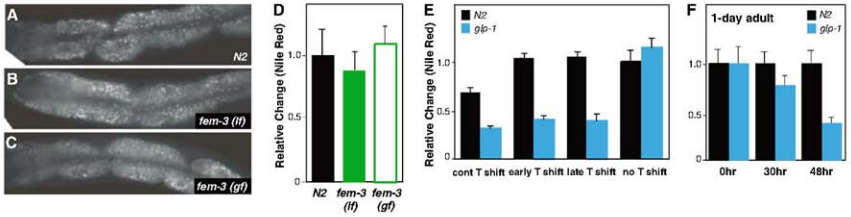


Fig. 2. Regulation of fat metabolism by GSC proliferation. **(A** to **D)** *fem-3(e2006)* loss-of-function mutants (*lf*) producing only oocytes or *fem-3(q20)* gain-of-function mutants (*gf*) generating only sperm showed the same fat storage as in the wild type ($P > 0.1$) ($n = 15$ for each). **(E)** Shift to 25°C during early or late larval development did not affect fat storage in the wild type ($P > 0.1$) but caused a 50% decrease in *glp-1(e2141)* mutants ($P < 0.0001$) ($n = 15$ for each genotype and treatment). **(F)** GSC arrest, caused by temperature shifting of 1-day-old *glp-1* adults, caused a decrease in fat storage (0 hours, $P > 0.5$; 30 hours, $P < 0.005$; 48 hours, $P < 0.0001$) ($n = 17$ for each genotype and treatment). **(G)** *lag-2(q420)* showed reduced fat ($P < 0.0001$) (*N2*, $n = 12$; *lag-2*, $n = 15$). **(H** to **K)** GSC overproliferation in the *glp-1(ar202)* gain-of-function mutant causes increased fat ($P < 0.0001$). In contrast, the loss-of-function mutant of *gld-1(q485)*, in which early-phase meiotic germ cells overproliferate, did not change fat storage ($P > 0.1$) (*N2*, $n = 12$; *glp-1*, $n = 15$; *gld-1*, $n = 13$).

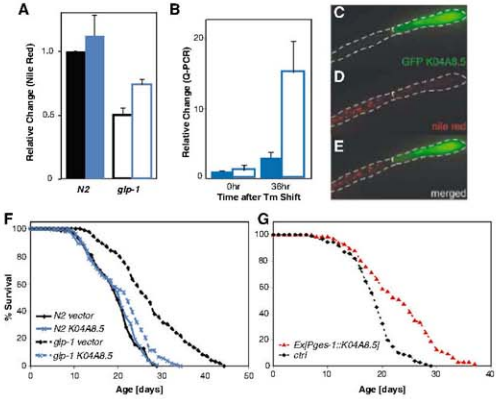


Fig. 3. A role for triglyceride lipase in lipid hydrolysis and longevity. **(A)** *K04A8.5* RNAi partially restored fat storage in *glp-1(e2141)* (open bars, $P < 0.001$) but had a marginal effect in the wild type (solid bars, 10% increase in fat storage, $P > 0.1$) ($n = 20$ for each genotype and RNAi feeding). Blue, *K04A8.5* RNAi; black, vector control. **(B)** *K04A8.5* expression was up-regulated in *glp-1* ($P < 0.0001$; solid bars, *N2*; open bars, *glp-1*). **(C** to **E)** Genetic mosaic analysis shows that the number and intensity of lipid droplets both decreased in the cell constitutively expressing *K04A8.5*, marked by GFP, relative to its sister cell. **(F)** *K04A8.5* RNAi had no effect on the life span of wild-type animals ($P > 0.01$) but suppressed the increased longevity of *glp-1* ($P < 0.0001$; 24% reduction in mean life span). **(G)** Constitutive expression of *K04A8.5* in the intestine extended life span ($P < 0.0001$; 24% mean life-span extension).

sized mitotic and meiotic germ cells and differentiated sperm. Despite a very different composition of the germ line, adult fat storage was decreased to a similar extent under all conditions (Fig. 2E). One process shared by all temperature shifts is germline stem cell (GSC) arrest (*16*), which could induce the decrease in fat storage. We therefore shifted temperature at 1 day of adulthood, after animals started to reproduce; this should affect adult GSCs but not the already proliferated germ line. By 30 hours at the restrictive temperature, fat storage in *glp-1* started to decrease (Fig. 2F). Within 48 hours, lipid accumulation in *glp-1* was reduced to an extent comparable to that seen with the developmental temperature shifts (Fig. 2F). This result suggests that GSCs regulate fat storage during adulthood.

The somatic distal tip cell forms the niche of GSCs. The Notch ligand LAG-2 expressed in the distal tip cell is required to maintain GSC identity (*17*). Like *glp-1* mutants, *lag-2(q420ts)* mutants (*18*) showed a 50% decrease in fat storage (Fig. 2G). *glp-1(ar202gf)* mutants with a hyperactive GLP-1, in which entry into meiosis is prevented and GSCs overproliferate (*19*), showed a factor of 1.7 fat increase (Fig. 2, H, I, and K), which suggests that a deficit of GSCs signals low fat storage and that GSC overproliferation signals high fat storage. No change in fat content was detected in *gld-1(q485)* mutants, in which early-phase meiotic germ cells reenter into the mitotic cell cycle and overproliferate (*20*) (Fig. 2, H, J, and K). Thus, once germ cells undergo differentiation, they lose the ability to modulate fat storage.

To understand the mechanisms by which GSCs regulate fat storage, we reduced the activities of 163 metabolic genes by RNA interference (RNAi) and screened for gene inactivations that increase fat storage in *glp-1* (table S1). Among 16 potential candidate genes identified, *K04A8.5* encoded a triglyceride lipase, which most strong-

ly affected fat storage. Inactivation of *K04A8.5* partially restored fat storage in *glp-1* but had marginal effect on the wild type (Fig. 3A). GSC arrest caused a marked increase in the transcriptional levels of *K04A8.5* (Fig. 3B), and a promoter–green fluorescent protein (GFP) reporter that was not detected under normal conditions became detectable in the *glp-1* gut at the restrictive temperature (fig. S4). High gene dosage of *K04A8.5* decreased fat storage in the wild type, and genetic mosaic animals showed that intestinal cells that constitutively express *K04A8.5* had fewer lipid droplets than did neighboring nontransgenic cells (Fig. 3, C to E). These results imply that this lipase acts in fat storage tissue rather than in endocrine cells or GSCs. Thus, the decrease in fat storage upon GSC arrest is induced by increased lipid hydrolysis via up-regulation of *K04A8.5*.

GSC arrest caused by *glp-1* loss of function resulted in extended life span (Fig. 3F and table S2) (*8*); *K04A8.5* RNAi suppressed this increased longevity but did not reduce wild-type life span (Fig. 3F and table S2). Therefore, up-regulation of this lipase gene mediates both lipid hydrolysis and longevity in GSC-arrested animals. Constitutive expression of *K04A8.5* specifically in the intestine led to life spans that were 24% longer than in control siblings (Fig. 3G and table S3). Thus, lipid hydrolysis in fat storage tissue prolongs life span, which connects the metabolic functions of adipose tissue to life-span control.

We investigated the signaling pathways regulating *K04A8.5* expression in the intestine. The forkhead transcription factor DAF-16 is translocated into nuclei in the intestine upon GSC arrest (*21*). To test whether *daf-16* is involved in regulation of fat storage by GSC proliferation, we inactivated *daf-16* by RNAi in wild-type and *glp-1* mutants and assayed fat storage. *daf-16* inactivation restored fat storage in *glp-1* but did not affect wild-type fat storage (Fig. 4A and fig. S5). *K04A8.5* up-regulation in *glp-1* was abolished in the absence of *daf-16* but was not altered in wild-type animals subjected to *daf-16* RNAi (Fig. 4B). Thus, upon GSC arrest, DAF-16 is activated in the intestine to promote lipid hydrolysis through induction of *K04A8.5* expression. External stresses such as heat shock and oxidative stress activate *daf-16* (*22*, *23*). After heat shock and paraquat treatment, the DAF-16 targets *hsp-16.1* and *ctl-2* were up-regulated but *K04A8.5* was not (fig. S6). These results suggest a specific regulation of *K04A8.5* by the signal from the germ line.

KRI-1, the human KRIT 1 homolog, and DAF-12, the nuclear hormone receptor, are both required for the intestinal nuclear localization of DAF-16 in GSC-arrested animals (*21*). These factors could act upstream of DAF-16 to sense signals from GSC and, in response, regulate lipid accumulation. Like *daf-16* RNAi, *kri-1* RNAi significantly reduced *K04A8.5* expression and increased the fat content in *glp-1* (Fig. 4, A and B, and fig. S5). In contrast, reducing *daf-12* function did not affect *K04A8.5* levels and caused a slight decrease in lipid accumulation in both wild-type and *glp-1* mutants (fig. S7). Therefore, GSC arrest promotes lipid hydrolysis in the intestine through activation of the *kri-1/daf-16* signaling pathway, but independently of *daf-12* lipophilic hormone signaling.

We examined *K04A8.5* expression in other long-lived animals, such as worms with reducing function in insulin receptor/*daf-2*. *daf-2* is crucial in regulation of fat metabolism during larval development (*24*). Therefore, we reduced *daf-2* function only at adulthood by RNAi feeding. Reducing *daf-2* activity at adulthood caused up-regulation of *K04A8.5* and decreased fat storage (Fig. 4, C and D, and fig. S8). Loss of the germ line and reduced *daf-2* signaling synergistically induced *K04A8.5* and decreased fat storage (Fig. 4, C and D, and fig. S8). We also found that *K04A8.5* RNAi partially suppressed the longevity of *daf-2* mutants (Fig. 4E and table S4). These results suggest that lipid hydrolysis is also connected to life-span control in the *daf-2* long-lived animals.

Our findings reveal an endocrine signaling axis from GSCs to fat storage tissue, with feedback from the fat storage to the longevity of the animal. Somatic stem cells are thought to mediate tissue regeneration after wounding, and such regeneration is also known to decline with aging. How the proliferation of adult stem cells is coupled to the requirement for replacement cells during normal and pathological aging may be related to the metabolic pathways we have discovered between germline stem cells and the longevity of *C. elegans*.

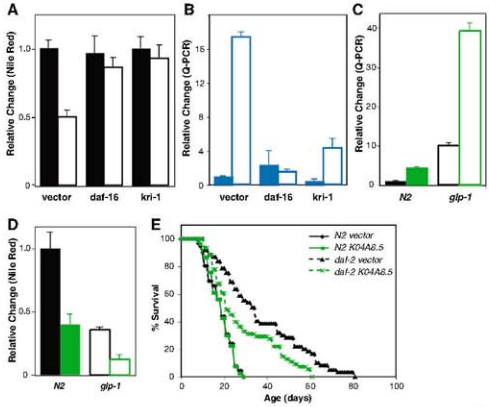


Fig. 4. Synergistic regulation of fat metabolism by GSC proliferation and insulin signaling. **(A)** Either *daf-16* or *kri-1* RNAi restored lipid accumulation in *glp-1* ($P < 0.0001$); neither of them affected fat storage in the wild type ($P > 0.05$) ($n = 17$ for each genotype and RNAi feeding). Solid bars, *N2*; open bars, *glp-1*. **(B)** *daf-16* and *kri-1* were required to up-regulate *K04A8.5* upon GSC arrest. *daf-16* or *kri-1* RNAi suppressed *K04A8.5* induction in *glp-1* ($P < 0.001$) ($n = 15$ for each genotype and RNAi feeding). Solid bars, *N2*; open bars, *glp-1*. **(C)** *K04A8.5* was induced in animals subjected to *daf-2* RNAi only at adulthood ($P < 0.0001$). This induction by *daf-2* RNAi was enhanced in the *glp-1* mutant ($P < 0.0001$). Green, *daf-2* RNAi; black, vector control. Solid bars, *N2*; open bars, *glp-1*. **(D)** Adult-specific *daf-2* RNAi decreased fat storage by 50% in the wild type ($P < 0.0001$). Loss of the germ line and reduction of *daf-2* activity were synergistic in reducing fat storage ($P < 0.0001$). Green, *daf-2* RNAi; black, vector control. Solid bars, *N2*; open bars, *glp-1*. **(E)** *K04A8.5* RNAi partially suppressed the longevity of *daf-2(e1370)* mutants ($P < 0.005$; 24% decrease in mean life span) but had no effect on the life span of the wild type ($P > 0.01$).

References and Notes

1. R. Zechner, J. G. Strauss, G. Haemmerle, A. Lass, R. Zimmermann, *Curr. Opin. Lipidol.* **16**, 333 (2005).
2. H. A. Tissenbaum, G. Ruvkun, *Genetics* **148**, 703 (1998).
3. D. J. Burks *et al.*, *Nature* **407**, 377 (2000).
4. M. C. Carr, *J. Clin. Endocrinol. Metab.* **88**, 2404 (2003).
5. M. Blüher, B. B. Kahn, C. R. Kahn, *Science* **299**, 572 (2003).
6. D. S. Hwangbo, B. Gershman, M. P. Tu, M. Palmer, M. Tatar, *Nature* **429**, 562 (2004).
7. M. E. Giannakou *et al.*, *Science* **305**, 361 (2004); published online 10 June 2004 (10.1126/science.1098219).
8. N. Arantes-Oliveira, J. Apfeld, A. Dillin, C. Kenyon, *Science* **295**, 502 (2002).
9. K. Ashrafi *et al.*, *Nature* **421**, 268 (2003).
10. M. J. Beanan, S. Strome, *Development* **116**, 755 (1992).
11. J. R. Priess, H. Schnabel, R. Schnabel, *Cell* **51**, 601 (1987).
12. H. Y. Mak, L. S. Nelson, M. Basson, C. D. Johnson, G. Ruvkun, *Nat. Genet.* **38**, 363 (2006).
13. S. Ogg *et al.*, *Nature* **389**, 994 (1997).
14. J. Hodgkin, *Genetics* **114**, 15 (1986).
15. M. K. Barton, T. B. Schedl, J. Kimble, *Genetics* **115**, 107 (1987).

16. J. E. Kimble, J. G. White, *Dev. Biol.* **81**, 208 (1981).
17. M. D. Wong, Z. Jin, T. Xie, *Annu. Rev. Genet.* **39**, 173 (2005).
18. S. T. Henderson, D. Gao, E. J. Lambie, J. Kimble, *Development* **120**, 2913 (1994).
19. A. S. Pepper, D. J. Killian, E. J. Hubbard, *Genetics* **163**, 115 (2003).
20. R. Francis, E. Maine, T. Schedl, *Genetics* **139**, 607 (1995).
21. J. R. Berman, C. Kenyon, *Cell* **124**, 1055 (2006).
22. S. S. Lee, S. Kennedy, A. C. Tolonen, G. Ruvkun, *Science* **300**, 644 (2003); published online 10 April 2003 (10.1126/science.1083614).
23. S. Wolff *et al.*, *Cell* **124**, 1039 (2006).
24. C. A. Wolkow, K. D. Kimura, M.-S. Lee, G. Ruvkun, *Science* **290**, 147 (2000).
25. We thank H. Mak, J. Dittman, and J. Avruch for critical reading of the manuscript; N. Ringstad for laser ablation techniques; V. Rottiers, A. Antebi, and the Caenorhabditis Genetics Center for providing strains; A. Fire for GFP vectors; and members of the Ruvkun lab for discussions. Supported by Life Sciences Research Foundation and Ellison Medical Foundation fellowships (M.C.W.), a Human Frontier Science Program fellowship (E.J.O.), and NIH grants 5R01AG016636 and 5R37AG14161 (G.R.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/957/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References

19 June 2008; accepted 8 September 2008
10.1126/science.1162011

Spontaneous Changes of Neocortical Code for Associative Memory During Consolidation

Kaori Takehara-Nishiuchi and Bruce L. McNaughton*

After learning, the medial prefrontal cortex (mPFC) gradually comes to modulate the expression of memories that initially depended on the hippocampus. We show that during this consolidation period, neural firing in the mPFC becomes selective for the acquired memories. After acquisition of memory associations, neuron populations in the mPFC of rats developed sustained activity during the interval between two paired stimuli, but reduced activity during the corresponding interval between two unpaired stimuli. These new patterns developed over a period of several weeks after learning, with and without continued conditioning trials. Thus, in agreement with a central tenet of consolidation theory, acquired associations initiate subsequent, gradual processes that result in lasting changes of the mPFC's code, without continued training.

The hippocampus is necessary for rapid association among elements of an event (*1*–*4*), and is initially also critical for retrieval of these associations; however, its necessity for retrieval is time-limited (*1*, *5*). In trace eyeblink conditioning, the medial prefrontal cortex (mPFC) becomes necessary for retrieval of associations as they become independent of the hippocampus (*6*), a process that requires intact *N*-methyl-D-aspartate (NMDA) receptor function in mPFC (*7*). If, as these results suggest, memory is gradually consolidated in a network encompassing mPFC, then mPFC neurons should become selective for learned associations with a similar time course.

We recorded from cells in the deep layers of prelimbic mPFC of rats, during a conditional associative learning task (table S1). Four rats were trained on a context-dependent association between a neutral tone [conditioned stimulus (CS)] and a mild shock to the eyelid [unconditioned stimulus (US)] (Fig. 1A). When the CS and US were paired in a fixed temporal pattern (Paired), the rats gradually expressed eyeblinks to the CS [conditioned response (CR), monitored by eyelid electromyogram] over ~10 days, and the frequency of CR expression was near asymptote throughout the recording sessions (Fig. 1B). In contrast, rats did not express CRs to the same CS when it was unpaired with the US (Pseudo) or presented alone (CS in box A and B). Selective neural activity for acquired associations was quantified with a discrimination function based on the difference in the neurons' responsiveness during the 500-ms interval after the CS (*8*) (Fig. 1C). The mPFC becomes necessary for retrieval 2 weeks or more after acquisition (*6*, *7*). Consistent with this time course, the selective neural activity for the association increased from the late stage of acquisition and reached a peak during the second week of overtraining (table S2A).

Arizona Research Laboratories, Division of Neural Systems, Memory, and Aging, University of Arizona, Tucson, AZ 85724–5115, USA.

*To whom correspondence should be addressed. E-mail: bruce@nsma.arizona.edu

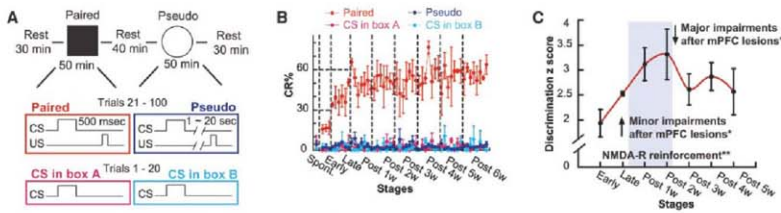


Fig. 1. Context-dependent acquisition of memory associations. **(A)** Rats were exposed to an environment in which an auditory CS was paired with eyelid stimulation US (Paired) and a separate environment in which the CS and US were unpaired (Pseudo), with each condition bordered by rest periods. During the first 20 trials in both conditions, the CS was delivered without the US. **(B)** Percentage of trials in which rats exhibited CRs increased in Paired, but not Pseudo, condition across weeks of learning (Early and Late) and overtraining (Post 1w to Post 6w; mean ± SEM from four rats; Spont.: spontaneous eyeblink frequency). **(C)** The mPFC becomes necessary for retrieval 2 weeks or more after acquisition [arrows (*6*)]. This process requires NMDA receptor function in the mPFC [shaded area (*7*)]. With a similar time course, neuron activity in the mPFC became selective for acquired associations (mean ± SEM from four rats). Selective activity was quantified on the basis of difference in firing-rate changes during the post-CS interval between Paired and Pseudo condition. Only the neurons that showed firing-rate differences between conditions were included.

We compared each neuron's firing rate during the CS or trace interval (interval between the CS and US) to its baseline activity (*8*). About 60% of neurons responded to the CS and/or trace interval during at least one condition throughout acquisition and overtraining. About 25% of neurons showed a different response pattern during Paired than during Pseudo and CS in box A (Behavioral context, Fig. 2B). Because they discriminated paired stimuli from unpaired stimuli, these neurons may encode acquired associations. Alternatively, they may have changed firing rates in conjunction with eyelid movements during Paired (i.e., CR), which were absent during Pseudo and CS in box A; however, this is unlikely, because firing rates of only ~5% of neurons were significantly correlated with CR amplitude or duration (fig.S1, A and B). Some neurons were sensitive to spatial context (i.e., the conditioning boxes) (~5%, Spatial context, Fig. 2C) or to whether the CS was presented alone or with the

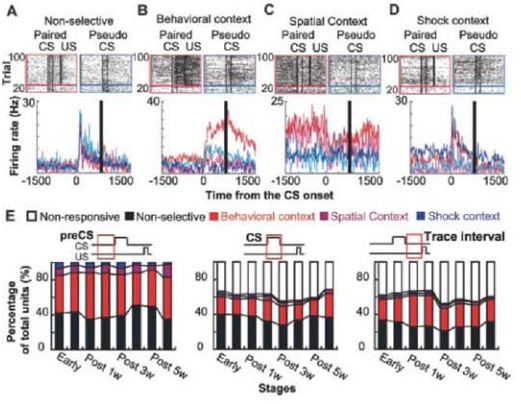


Fig. 2. Response categories of activity in single neurons. (**A** to **D**) Although some neurons showed the same stimulus-related neural firing patterns regardless of context (**A**), others were found with firing rates that depended on behavioral context [rates in Paired were different from rates in Pseudo, and from rates in CS in box A condition (B)], spatial context [rates in Paired were different from rates in Pseudo, but not from rates in CS in box A condition (C)], or shock context [rates in Paired were different from rates in CS in box A, but not from rates in Pseudo condition (D)]. Raster plots and peristimulus time histograms show firing rates during a ±1500-ms period around the CS onset (1-ms bins, smoothed with 50-ms hamming window; pink: CS in box A; red: Paired; cyan: CS in box B; blue: Pseudo). The black filled bar indicates a blackout period because of the US artifact. (**E**) There were no changes in the proportion of cells in each category across stages of learning and overtraining. This was true when examining the activity during the pre-CS period (left), the CS (middle), or the trace interval (right).

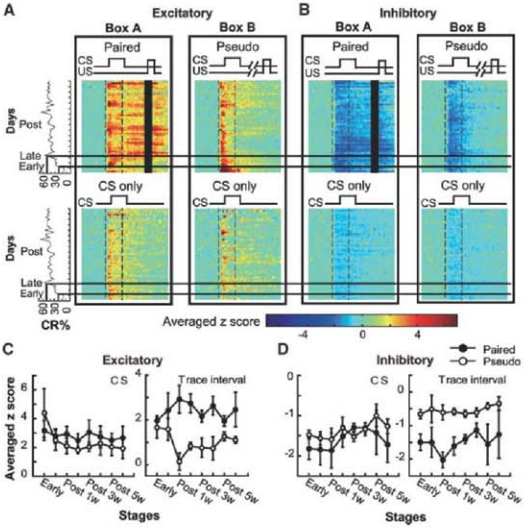


Fig. 3. Plasticity in firing-rate patterns of mPFC neural populations. (**A**) *Z*-score values for neuron responses to the CS and trace interval (i.e., change from baseline during a 1800-ms window around the CS, truncated to 8 and −8 for illustration purposes) were averaged across all excitatory neurons, and this measure of population activity is plotted over time within trials (*x* axis) across days (*y* axis, ascending from bottom to top; averaged percent CRs are plotted to the left). During acquisition and overtraining, mPFC neurons exhibited sustained excitatory activity in the trace interval during Paired condition (left). During overtraining, this sustained activity reduced during Pseudo condition (right), and was absent in trials with the CS alone (lower panels). (**B**) Medial prefrontal neurons with inhibitory responses exhibited a sustained response in the trace interval specifically during Paired. (**C**) Neural activity in the population of cells with excitatory responses (averaged *z* scores) did not differ between Paired and Pseudo conditions during the CS (left), but were different between conditions during the trace interval after CRs were acquired (mean ± SEM from four rats). (**D**) Neurons exhibiting inhibitory responses differentiated between Paired and Pseudo trace interval before CRs were fully acquired.

US (~5%, Shock context, Fig. 2D). Others showed the same response pattern in all four conditions (~35%, Non-selective, Fig. 2A). About 40% of neurons changed their baseline rates during Paired compared to Pseudo, suggesting that their activity was selective for the context per se or for different background behaviors in the two contexts. No changes in the ensemble pattern of these response categories were observed across two stages of acquisition and six stages of overtraining (Fig. 2E).

To examine how the activity patterns of single neurons generalized to the population and how the population response patterns changed during acquisition and overtraining, we normalized the activity of each neuron and used the averages within the same response type (excitatory or inhibitory, fig. S2) (*8*) as a measure of population response pattern (Fig. 3). At the beginning of acquisition (Early), the excitatory response pattern during Paired was similar to the pattern during Pseudo; however, as the rats acquired CRs, response patterns in the trace interval during Paired became separated from those in the corresponding interval during Pseudo (Fig. 3, A and C, right, and table S2B). During Paired, a sustained response to the trace interval was observed from the early stage of acquisition and was maintained for a 6-week period of overtraining (Fig. 3C, right, and table S2C). In contrast, during Pseudo, the response to the corresponding interval gradually weakened during overtraining (Fig. 3C, right, and table S2D). The excitatory response to the CS was similar between Paired and Pseudo and gradually weakened during overtraining (Fig. 3C, left, and table S2E). Similarly, the response of neurons that exhibited inhibition in the interval between the stimuli was stronger during Paired than during Pseudo from the beginning of acquisition to the final week of overtraining (Fig. 3, B and D, right, and table S2F). The inhibitory response to the CS was similar between Paired and Pseudo during acquisition and overtraining (Fig. 3, B and D, left, and table S2G). Population excitatory and inhibitory responses were weak and transient to the CS when the CS was presented alone throughout all of the stages (Fig. 3, A and B).

To clarify whether these new firing rate patterns resulted from repetitive overtraining or a spontaneous neural process subsequent to the initial acquisition of the memory, we recorded from three additional rats (rats 5 to 7) during reconditioning (Retention) after a 6-week, training-free "consolidation period" (Fig. 4B). Although the percentage of CRs was reduced in the first retention session, it was significantly higher during Paired than during Pseudo (table S2H) and spontaneous eyeblink frequency (table S2I). Moreover, the rats relearned the CR faster than during the initial conditioning: The number of sessions required to reach 60% CRs was 4.7 ± 0.4, whereas during initial acquisition, 10.7 ± 0.8 sessions were required (table S2J). The mPFC population showed a sustained excitatory response to the CS during Paired, but only a transient response to the CS during Pseudo (Fig. 4D, top). These activity patterns were similar to the patterns during the final (sixth) week of overtraining (Post 6w, Fig. 4C, top right), but not during learning (Late, Fig. 4C, top left). The population inhibitory response during Paired was larger than it was during Pseudo (Fig. 4D, bottom), in a manner similar to that of the patterns during Late (Fig. 4C, bottom left) and Post 6w (Fig. 4C, bottom right). Further control experiments (fig. S3) confirmed that differential activity between the paired stimuli and the unpaired stimuli could not be attributed to the fixed order of conditions or to instability of neural activity.

In addition to changes in firing rate, information about memory associations may also be encoded in relative spike-timing patterns of multiple neurons. The template-matching method (*9*, *10*) showed that population spike-timing patterns gradually became selective for associations as a consequence of overtraining, but not simply as a consequence of the passage of time (figs. S4 and S5).

Consolidation of memory is presumed to involve gradual reorganization of cortical networks (*2*, *6*, *11*–*16*). Our findings suggest that this consolidation process may be directly reflected in the activity pattern of the mPFC neurons. As the rats acquired the conditional association, the firing rates of some mPFC neurons became selective for the acquired associations, by exhibiting sustained activity during the interval between the paired stimuli, and corre-

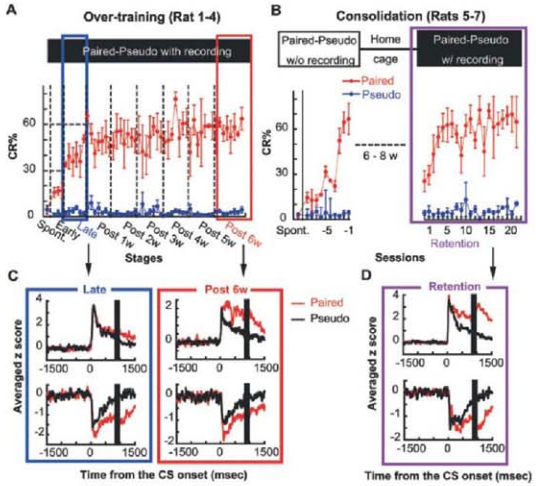


Fig. 4. Change of neural activity patterns with time, without overtraining. **(A)** Mean percentage of trials with CRs increased specifically during Paired condition in the overtraining group (mean ± SEM from four rats). **(B)** The consolidation group maintained higher percent CRs in Paired as compared with Pseudo condition after a 6- to 8-week period without exposure to the conditioning environments (mean ± SEM from three rats; Spont.: spontaneous eyeblink frequency). **(C)** In the overtraining group, excitatory firing-rate patterns differentiated between Paired and Pseudo conditions during the sixth week of overtraining (Post 6w, right), but not during the late stage of acquisition (Late, left). **(D)** In the consolidation group, population firing-rate patterns during reconditioning were similar to those observed after overtraining. Plots show averaged firing-rate patterns during ±1500-ms time windows around the CS onset, binned at 1 ms, for neurons with excitatory responses (top) and inhibitory responses (bottom).

spondingly reduced activity during the interval between the unpaired stimuli. Considering that a proportion of single neurons with selective responses to the paired stimuli did not increase across sessions of acquisition or overtraining, a specific set of mPFC neurons encodes the difference between paired and unpaired stimuli, and changes in the magnitude of their differential activity, rather than changes in the number of differentiating neurons, are responsible for the gradual development of activity selective for memory associations. In addition, some neurons changed their baseline firing rate depending on the context (Fig. 2, C to E), indicating that acquired associations modify not only neural responses to presented stimuli, but also responses to the context.

Population firing-rate activity selective for memory associations was observed from the late stage of acquisition to 6 weeks of overtraining. Similar patterns developed after a 6-week period without any conditioning. This slow time course of firing-rate changes agrees with a previously observed time window in which the mPFC becomes important for retrieval (*6*). One possible mechanism of these lasting changes may be that the weights of intra- and inter-area connections in the mPFC are gradually redistributed by multiple rounds of synaptic modification (*7*). Such synaptic modification might be triggered by the replay of task-related neural activity patterns during sleep (*17*). In contrast, hippocampal neurons change firing-rate patterns at the onset of learning (*18*, *19*). Therefore, the hippocampus may direct the mPFC to refine its memory-related neural activity, as proposed in standard consolidation theory (*2*, *15*, *16*, *20*, *21*).

Memories of associations between various elements of an event are presumed to be distributed over wide areas of cortex. The hippocampus may store a memory index for a unique array of neocortical modules representing each experiential event (*22*). Direct cortico-cortical connections that are gradually established during consolidation may render the hippocampal index codes for the original memory no longer necessary; alternatively, for certain forms of memory, the mPFC might take over the linking function from the hippocampus by storing a similar, but perhaps more efficient, index code (*23*). The observed selective activity for context-dependent, cross-modal associations in the mPFC is a critical prerequisite for the mPFC to function as a storage site of memory indices for consolidated memory. The role of the mPFC in the retrieval of consolidated memory may thus parallel the putative role of the hippocampus in retrieving recently formed memories, by completing the intercortical neural pattern from a partial cue. This role may be related to the involvement of human mPFC in memory retrieval (*23*, *24*) or retrieval judgment, such as a feeling of knowing (*25*–*28*).

References and Notes

1. W. B. Scoville, B. Milner, *J. Neurol. Neurosurg. Psychiatry* **20**, 11 (1957).
2. D. Marr, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **262**, 23 (1971).
3. M. Moss, H. Mahut, S. Zola-Morgan, *J. Neurosci.* **1**, 227 (1981).
4. B. L. McNaughton, R. G. M. Morris, *Trends Neurosci.* **10**, 408 (1987).
5. L. R. Squire, *Psychol. Rev.* **99**, 195 (1992).
6. K. Takehara, S. Kawahara, Y. Kirino, *J. Neurosci.* **23**, 9897 (2003).
7. K. Takehara-Nishiuchi, K. Nakao, S. Kawahara, N. Matsuki, Y. Kirino, *J. Neurosci.* **26**, 5049 (2006).
8. Materials and methods are available as supporting material on *Science* Online.
9. K. Louie, M. A. Wilson, *Neuron* **29**, 145 (2001).
10. M. Tatsuno, P. Lipa, B. L. McNaughton, *J. Neurosci.* **26**, 10727 (2006).
11. B. Bontempi, C. Laurent-Demir, C. Destrade, R. Jaffard, *Nature* **400**, 671 (1999).
12. J. H. Freeman Jr., M. Gabriel, *Neurobiol. Learn. Mem.* **72**, 259 (1999).
13. P. W. Frankland, B. Bontempi, L. E. Talton, L. Kaczmarek, A. J. Silva, *Science* **304**, 881 (2004).
14. T. Maviel, T. P. Durkin, F. Menzaghi, B. Bontempi, *Science* **305**, 96 (2004).
15. J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, *Psychol. Rev.* **102**, 419 (1995).
16. L. R. Squire, P. Alvarez, *Curr. Opin. Neurobiol.* **5**, 169 (1995).
17. D. R. Euston, M. Tatsuno, B. L. McNaughton, *Science* **318**, 1147 (2007).
18. M. D. McEchron, J. F. Disterhoft, *Hippocampus* **9**, 385 (1999).
19. S. Wirth *et al.*, *Science* **300**, 1578 (2003).
20. G. Buzsaki, *Neuroscience* **31**, 551 (1989).
21. K. A. Paller, *Memory* **5**, 73 (1997).
22. T. J. Teyler, P. DiScenna, *Behav. Neurosci.* **100**, 147 (1986).
23. A. Takashima *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 756 (2006).
24. S. Gais *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18778 (2007).
25. M. Moscovitch, B. Melo, *Neuropsychologia* **35**, 1017 (1997).
26. H. Kikyo, K. Ohki, Y. Miyashita, *Neuron* **36**, 177 (2002).
27. D. M. Schnyer *et al.*, *Neuropsychologia* **42**, 957 (2004).
28. D. M. Schnyer, L. Nicholls, M. Verfaellie, *J. Cogn. Neurosci.* **17**, 832 (2005).

29. This work was supported by NIH grants NS020331 and MH064823 (B.L.M.) and the Human Frontier Science Program (K.T.-N.). We thank M. Tatsuno and S. Cowen for helping with construction of the experimental apparatus and programming of the code for data analysis; J. Meltzer, J. Wang, H. DeFour, and A. Mosley for help with data collection and analysis; K. Bohne for constructing hyperdrives; and N. Insel for critical reading of the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/960/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
Tables S1 to S6
References

3 June 2008; accepted 26 September 2008
10.1126/science.1161299

Promoting Axon Regeneration in the Adult CNS by Modulation of the PTEN/mTOR Pathway

Kevin Kyungsuk Park,* Kai Liu,* Yang Hu,* Patrice D. Smith,* Chen Wang, Bin Cai, Bengang Xu, Lauren Connolly, Ioannis Kramvis, Mustafa Sahin, Zhigang He†

The failure of axons to regenerate is a major obstacle for functional recovery after central nervous system (CNS) injury. Removing extracellular inhibitory molecules results in limited axon regeneration in vivo. To test for the role of intrinsic impediments to axon regrowth, we analyzed cell growth control genes using a virus-assisted in vivo conditional knockout approach. Deletion of PTEN (phosphatase and tensin homolog), a negative regulator of the mammalian target of rapamycin (mTOR) pathway, in adult retinal ganglion cells (RGCs) promotes robust axon regeneration after optic nerve injury. In wild-type adult mice, the mTOR activity was suppressed and new protein synthesis was impaired in axotomized RGCs, which may contribute to the regeneration failure. Reactivating this pathway by conditional knockout of tuberous sclerosis complex 1, another negative regulator of the mTOR pathway, also leads to axon regeneration. Thus, our results suggest the manipulation of intrinsic growth control pathways as a therapeutic approach to promote axon regeneration after CNS injury.

Axons do not regenerate after injury in the adult mammalian central nervous system (CNS), a phenomenon attributed to two properties of the adult CNS, the inhibitory extrinsic environment and a diminished intrinsic regenerative capacity of mature CNS neurons (*1*–*4*). Neutralization of the extracellular molecules identified as axon regrowth inhibitors allows only a limited degree of axon regeneration in vivo (*5*–*7*). Therefore, intrinsic mechanisms are likely to be important in controlling the process of axon regeneration. A hint about possible mechanisms of neuronal regenerative ability comes from the evolutionarily conserved molec-

ular pathways that control cellular growth and size. For most cell types, specific mechanisms are necessary to prevent cellular overgrowth upon the completion of development (*8*). Because many of these molecules are often expressed in postmitotic mature neurons, we hypothesized that they may contribute to the diminished regenerative ability in adult CNS neurons.

To circumvent the problem that germline knockout of individual cell growth control genes often results in compromised viability in mice, we designed a strategy based on intravitreal injection of adeno-associated viruses expressing Cre (AAV-Cre) in adult mice. This procedure resulted in the expression of Cre in more than 90% of retinal ganglion cells (RGCs) and few other non-RGC cells, as indicated in two reporter lines (fig. S1, A and B). We thus injected AAV-Cre into the vitreous body of different adult floxed mice, including $Rb^{f/f}$ (*9*), $P53^{f/f}$ (*9*), $Smad4^{f/f}$ (*10*), $Dicer^{f/f}$ (*11*), $LKB1^{f/f}$ (*12*), and $PTEN^{f/f}$ (*13*). At 2 weeks after AAV injection (fig. S1C), these animals were subjected to an optic nerve crush procedure (*14*). Axon regeneration was assayed by examining axonal fibers labeled with the anterograde tracer, cholera toxin β (CTB), in the optic nerve sections across the lesion site. Neuronal survival was also estimated by whole-mount staining of the retina with immunostaining with an antibody to β-III tubulin (TUJ1) or by prelabeling with FluoroGold injected to the superior colliculi.

Among the different mouse lines examined, those with a PTEN (phosphatase and tensin homolog) deletion showed the largest effects on both neuronal survival and axon regeneration. In all $PTEN^{f/f}$ conditional mutants injected with AAV-Cre, but not with control AAV-GFP (green fluorescent protein), RGCs displayed a significant increase in survival as indicated by TUJ1 staining (Fig. 1F and fig. S1D). Similar extents of neuronal survival were observed by prelabeling RGCs with FluoroGold (fig. S2). In addition, robust long-distance axon regeneration (Fig. 1, A to E, and fig. S3, A and B) was observed at 14 days after injury. We repeated the AAV-Cre experiments in another set of $PTEN^{f/f}$ mice, as well as AAV-GFP injection as controls, and observed similar results. Quantification showed that ~45% of PTEN-deleted RGCs survived 2 weeks after injury, in comparison to ~20% in controls. Of the surviving RGCs, ~8 to 10% showed regeneration of their lesioned axons beyond 0.5 mm distal to the lesion epicenter; some extended beyond 3 mm (Fig. 1, A and E, and fig. S3, A and B). These fibers continued to regenerate along the optic nerve over time. At 4 weeks after injury, some regenerating fibers extended to the area of the optic chiasm (Fig. 1, B, D, and E, and fig. S3, C and D). Among the other mouse lines tested, only p53-deleted RGCs showed increased neuronal survival but no axon regeneration (fig. S1, D and E), consistent with the notion that inducing neu-

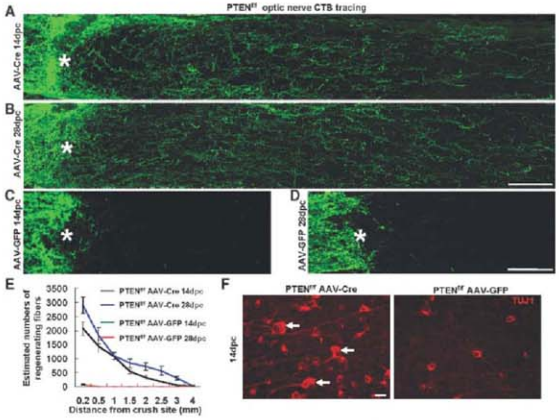


Fig. 1. PTEN deletion promotes RGC axon regeneration. (**A** to **D**) Confocal images of optic nerves showing CTB-labeled fibers around the lesion sites at 14 days [(A) and (C)] or 28 days [(B) and (D)] after injury from $PTEN^{f/f}$ mice injected with AAV-Cre [(A) and (B)] or AAV-GFP [(C) and (D)]. Scale bar, 100 μm. (**E**) Quantification of regenerating fibers at different distances distal to the lesion sites. At least five different sections (every fourth section) from each animal were quantified. At both time points, there were significant differences between control and PTEN-deleted mice groups at every distance measured by analysis of variance with Bonferroni's post-test [$P < 0.05$ for both 14 dpc ($n = 7$) and 28 dpc ($n = 3$)]. (**F**) Fluorescent photomicrographs of retinal whole-mounts showing surviving TUJ1$^+$ RGCs at 14 days after injury. Arrows, RGCs with enlarged soma; *, crush site; scale bar, 20 μm.

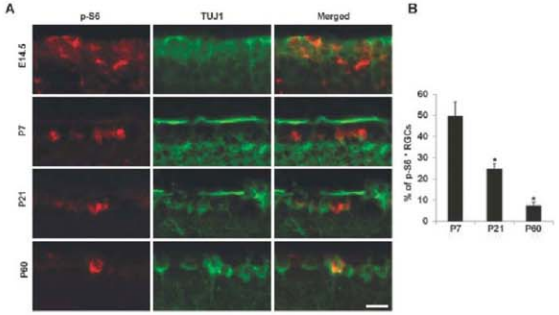


Fig. 2. Development-dependent decline of p-S6 signals in RGCs. Representative images (**A**) and quantification (**B**) of immunohistochemical analysis for p-S6 or TUJ1 immunoreactivity of the ganglion cell layer (GCL) of mouse retinas at different ages. Scale bar, 20 μm. Data are presented as mean percentages of p-S6$^+$TUJ1$^+$ cells among total TUJ1$^+$ cells in the GCL of each retina. There is a significant difference in percentages of P21 and P60 when compared to that of P7. *, $P < 0.01$, Student's t test. Cell counts were performed on at least four nonconsecutive sections, from three mice per age group.

F. M. Kirby Neurobiology Center, Children's Hospital, and Department of Neurology, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: zhigang.he@childrens.harvard.edu

ronal survival is not sufficient to allow axon regeneration (*15*). Thus, PTEN deletion likely acts upon intrinsic mechanisms other than survival to promote axon regeneration after injury.

We next examined the time course of axon elongation of PTEN-deleted RGCs after injury. At the lesion site, a glial response occurred 1 to 3 days after crushing, as indicated by the up-regulation of chondroitin sulfate proteoglycan (CSPG) expression (fig. S4, A and B) and macrophage infiltration (fig. S5). However, glial fibrillary acidic protein (GFAP)–positive astrocytes are largely excluded from the lesion sites, and CSPG signals returned to low levels by 7 days after injury (fig. S4C). At 1 day post-crush (dpc), injured optic nerve fibers terminate at the proximal end of the crush site in PTEN$^{f/f}$ mice injected with either AAV-GFP or AAV-Cre (fig. S4A). However, at 3 dpc, axonal sprouts from PTEN-deficient RGCs began to penetrate the CSPG-enriched lesion site (fig. S4, B and D, and fig. S5A), and some fibers could be seen beyond the lesion sites 7 dpc (figs. S4, C and E, and fig. S5B). In contrast, minimum axonal sprouts were seen in control animals at these stages (fig. S4, A to E, and fig. S5, A and B). Electron microscopic analysis confirmed that in wild-type mice, degenerating RGC axons, myelin debris, and macrophages occupied injury sites, and few regenerating fibers were visible (fig. S5, C and E). However, when PTEN was deleted in RGCs, regenerative axonal sprouts, often appearing as bundles, were found both within and distal to the lesion site soon after injury (fig. S5, D and F). These results indicate that PTEN deletion indeed enabled axons to overcome inhibition at the lesion site and to regenerate soon after injury.

PTEN deletion activates the PI3K/mTOR (mammalian target of rapamycin) pathway, which controls cell growth and size by regulating cap-dependent protein translation initiation (*16*, *17*) (fig. S6). Two of the well-studied targets of the mTOR kinase are ribosomal S6 kinase 1 (S6K1) (which in turn phosphorylates ribosomal protein S6) and the eukaryotic initiation factor 4E (eIF4E)-binding protein 4E-BP1. We used anti-

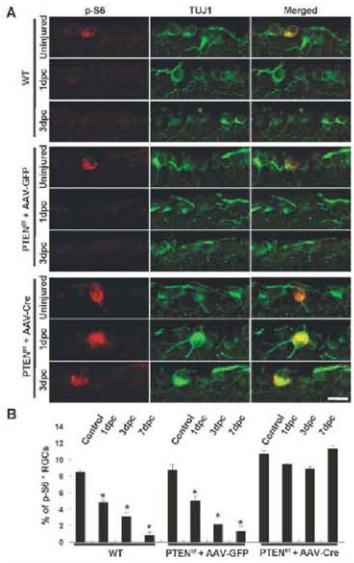


Fig. 3. Axotomy reduces p-S6 levels in RGCs in control but not in PTEN-deleted RGCs. **(A)** Immunofluorescence analysis with antibodies to p-S6 or TUJ1 of the retinal sections from wild-type or PTEN$^{f/f}$ mice injected with AAV-GFP or AAV-Cre in different animal groups. Scale bar, 20 μm. **(B)** Quantification of p-S6$^+$ RGCs. Data are presented as mean percentages of p-S6$^+$TUJ1$^+$ cells among total TUJ1$^+$ cells in the GCL of each retina. Cell counts were performed on at least four nonconsecutive sections, from three mice per group. *, $P < 0.01$ by Dunnett's test. Comparisons were made against the uninjured control retinas.

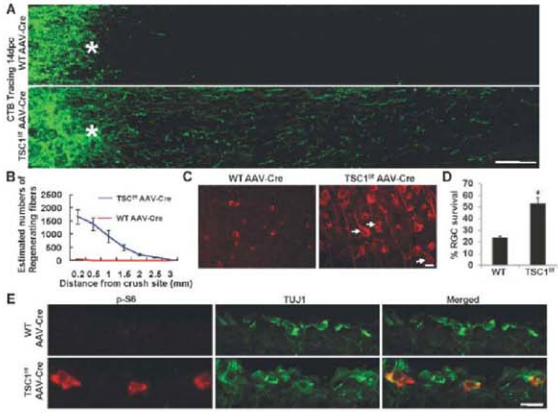


Fig. 4. TSC1 deletion promotes RGC survival and axon regeneration after optic nerve crush. **(A)** Confocal images of optic nerves spanning lesion sites from wild-type or TSC1$^{f/f}$ mice injected with AAV-Cre at 14 days after crush. Scale bar, 100 μm. **(B)** Quantification of regenerating fibers in different groups. There is a significant difference between control and TSC1-deleted groups by Student's *t* test at each distance ($P < 0.05$). **(C)** Fluorescent photomicrographs of retinal whole-mounts showing characteristic surviving TUJ1$^+$ RGCs. Arrows, RGCs with enlarged soma; *, crush site; scale bar, 20 μm. **(D)** RGC survival at 14 dpc, expressed as a percentage of the total number of TUJ1$^+$ RGCs in the uninjured retina. *, $P < 0.05$, Student's *t* test. **(E)** Immunofluorescence analysis of the retinal sections from AAV-Cre–injected wild-type or TSC1$^{f/f}$ mice at 14 days after injury with antibodies to p-S6 or TUJ1. Scale bar, 20 μm.

bodies to phospho-S6 (p-S6) to monitor the activity of the mTOR pathway in the RGCs. p-S6 staining revealed a development-dependent decline in the percentage of p-S6–positive RGCs (Fig. 2, A and B). Strong p-S6 signals can be seen in most embryonic neurons but are diminished in 90% of adult RGCs (Fig. 2, A and B), suggesting that mTOR signaling is down-regulated in the majority of adult RGCs, with only a small subset retaining considerable mTOR activity.

We next investigated whether axon injury alters the activity of the mTOR pathway and subsequent protein synthesis in adult RGCs. Upon experiencing stresses such as hypoxia, cells suppress the mTOR signaling (*18*–*20*), an evolutionally conserved stress response proposed to maintain energy homeostasis for survival. Thus, we postulated that the axotomy-triggered stress response may reduce global protein translation in injured neurons. To test this, we estimated the rate of new protein synthesis in purified RGCs from control and injured rats (figs. S7 and S8). Control and injured RGCs were incubated with ^{35}S-methanine/cysteine, and the results showed a significant decrease in new protein synthesis in axotomized RGCs (fig. S8).

We then studied whether axon injury suppressed the mTOR activity by immunohistochemistry of retinal sections with antibodies to p-S6. Axotomy almost completely abolished the remaining p-S6 signals in adult RGCs at all post-injury time points examined (1, 3, and 7 days) in control mice (Fig. 3, A and B), indicating that axotomy triggers a rapid and sustained down-regulation of the mTOR activity in all neurons. In cells under hypoxic stress, up-regulation of Redd1/2 (regulated in development and DNA damage responses 1/2) is involved in the inhibition of mTOR (*18*–*20*). However, Redd1/2 mRNA levels were not altered in axotomized RGCs (fig. S9), suggesting that mechanisms other than Redd1/2 mediate the suppression of the mTOR pathway after axotomy.

We next assessed whether PTEN deletion affects the mTOR activity in the adult RGCs both before and after injury. AAV-Cre injection did not significantly increase the percentage of RGCs stained with antibodies to p-S6 in uninjured PTEN$^{f/f}$ mice, which may be due to the short interval between AAV-Cre injection and histological analysis (2 weeks). However, after optic nerve crush, similar percentages of p-S6–positive RGCs remained at 1, 3, or 7 days after injury in the AAV-Cre–injected PTEN$^{f/f}$ mice (Fig. 3, A and B). Thus, despite experiencing a stress response to axotomy, these PTEN-deleted neurons were able to maintain mTOR activity levels similar to those of uninjured wild-type neurons. Notably, the percentage of regenerating fibers (8 to 10%) was similar to that of p-S6–positive axotomized RGCs (Fig. 1E).

To further assess the relationship between p-S6–positive RGCs and RGCs that regenerated axons, we injected biotinylated dextran-amine (BDA), an axonal tracer, to the optic nerve region ~2 to 2.5 mm distal to the lesion site. This method resulted in few BDA-labeled cells in the retinas in wild-type mice at 14 days after injury (fig. S10A), consistent with the lack of regeneration of in wild-type animals. In PTEN-deleted animals, this same procedure labeled, on average, a total of 68 RGCs per retina, and 88% of these labeled RGCs were positively stained with antibodies to p-S6 (fig. S10). This result showed that the majority of regenerating fibers are from those p-S6–positive RGCs.

We next assessed the necessity of mTOR activation for the regeneration effects of PTEN-deletion. Administration of rapamycin, an mTOR inhibitor, markedly reduced p-S6 signals in the RGCs (fig. S11, E and F) and largely neutralized the survival and regeneration observed in PTEN$^{f/f}$ mice injected with AAV-Cre (fig. S11, A to D). Although residual axon regeneration observed in the PTEN-deleted mice treated with rapamycin may reflect possible involvements of other pathways downstream of PTEN, these results demonstrated a major role played by the mTOR activation in the effects of PTEN deletion.

To further examine whether activation of the mTOR pathway is sufficient to promote axon regeneration, we performed the optic nerve injury model in TSC1$^{f/f}$ (tuberous sclerosis complex 1) conditional knockout mice (*21*). TSC1/2 form a protein complex that negatively regulates mTOR signaling, and loss of either TSC1 or TSC2 leads to constitutive activation of the mTOR pathway (*22*–*24*) (fig. S6). As expected, in AAV-Cre–injected TSC1$^{f/f}$ mice, strong p-S6 signals were observed in axotomized RGCs (Fig. 4E), and RGC survival was enhanced after injury (Fig. 4, C and D). More important, considerable axon regeneration (Fig. 4, A and B) was observed in TSC1-deleted but not in wild-type mice injected with AAV-Cre. The extent of axon regeneration in TSC1-deleted mice was slightly weaker than that induced by PTEN deletion (comparing Fig. 1, A and E, to Fig. 4, A and B). Together with the observations that some p-S6–negative RGCs were labeled with BDA applied to the optic nerve distal to the lesion (fig. S10) and that the residual regeneration was observed in PTEN-deleted mice treated with rapamycin (fig. S11), these results suggest that changes in other downstream targets of PTEN, such as Akt and GSK-3 (*25*, *26*), may also contribute to regenerative growth. Nonetheless, these data together indicate that activation of the mTOR pathway is sufficient to promote both RGC survival and axon regeneration.

Our findings reveal an important role for the regulation of the mTOR pathway and protein translation in determining the intrinsic axon-regrowth responsiveness of injured CNS neurons. This pathway is profoundly suppressed in axotomized adult neurons, which may limit new protein synthesis required for sustained axon regeneration. Reactivating the mTOR pathway by silencing either PTEN or TSC1 in adult neurons induces extensive axon regeneration, suggesting that retaining active protein synthesis in axotomized mature neurons is sufficient to initiate a neuronal regenerative program for axon regrowth. In light of the availability of specific chemical inhibitors of PTEN (*27*), these and other strategies may potentially be used transiently after CNS injury to prevent the down-regulation of protein synthesis and promote axon regeneration and functional recovery.

References and Notes

1. M. E. Schwab, D. Bartholdi, *Physiol. Rev.* **76**, 319 (1996).
2. J. L. Goldberg, M. P. Klassen, Y. Hua, B. A. Barres, *Science* **296**, 1860 (2002).
3. M. T. Filbin, *Philos. Trans. R. Soc. London B Biol. Sci.* **361**, 1565 (2006).
4. M. T. Filbin, J. Silver, *Exp. Neurol.* **209**, 294 (2008).
5. L. C. Case, M. Tessier-Lavigne, *Curr. Biol.* **15**, R749 (2005).
6. G. Yiu, Z. He, *Nat. Rev. Neurosci.* **7**, 617 (2006).
7. N. Y. Harel, S. M. Strittmatter, *Nat. Rev. Neurosci.* **7**, 603 (2006).
8. R. A. Weinberg, in *The Biology of Cancer*, R. A. Weinberg, Ed. (Garland Science, 2007), pp. 209–253.
9. S. Marino, M. Vooijs, H. van Der Gulden, J. Jonkers, A. Berns, *Genes Dev.* **14**, 994 (2000).
10. X. Yang, C. Li, P. L. Herrera, C. Deng, *Genesis* **32**, 80 (2002).
11. B. D. Harfe, M. T. McManus, J. H. Mansfield, E. Hornstein, C. J. Tabin, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10898 (2005).
12. N. Bardeesy *et al.*, *Nature* **419**, 162 (2002).
13. M. Groszer *et al.*, *Science* **294**, 2186 (2001).
14. S. Leon, Y. Yin, J. Nguyen, N. Irwin, L. L. Benowitz, *J. Neurosci.* **20**, 4615 (2000).
15. J. L. Goldberg *et al.*, *Neuron* **33**, 689 (2002).
16. D. A. Guertin, D. M. Sabatini, *Cancer Cell* **12**, 9 (2007).
17. J. Luo, B. D. Manning, L. C. Cantley, *Cancer Cell* **4**, 257 (2003).
18. J. Brugarolas *et al.*, *Genes Dev.* **18**, 2893 (2004).
19. M. N. Corradetti, K. Inoki, K. L. Guan, *J. Biol. Chem.* **280**, 9769 (2005).
20. J. H. Reiling, E. Hafen, *Genes Dev.* **18**, 2879 (2004).
21. L. Meikle *et al.*, *J. Neurosci.* **27**, 5546 (2007).
22. A. Garami *et al.*, *Mol. Cell* **11**, 1457 (2003).
23. K. Inoki, T. Zhu, K. L. Guan, *Cell* **115**, 577 (2003).
24. A. R. Tee *et al.*, *Curr. Biol.* **13**, 1259 (2003).
25. W. Y. Kim *et al.*, *Neuron* **52**, 981 (2006).
26. K. Namekawa *et al.*, *J. Neurosci.* **20**, 2875 (2000).
27. E. Rosivatz *et al.*, *ACS Chem. Biol.* **1**, 780 (2006).

28. We thank H. Hasegawa and F. Wang for providing Rosa-stop-PLAP mice, D. Kwiatkowski for proving TSC1$^{f/f}$ mice, C. Deng for Smad4$^{f/f}$ mice, M. Ericsson and L. Bu for microscopic analysis, and L. Benowitz, M. Greenberg, T. Hensch and J. Kanter for reading the manuscript. This study was supported by grants from Craig H. Neilson Foundation (to K.K.P.); National Institute of Neurological Disorders and Stroke (NINDS), Wings for Life, and Adelson Foundation (to Z.H.); Ruth L. Kirschstein National Research Service Award (NRSA) (to Y.H.); Canadian Institute of Health Research (to P.D.S.); and NINDS and Tuberous Sclerosis Alliance (to M.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/963/DC1
Materials and Methods
Figs. S1 to S13
References

9 June 2008; accepted 12 September 2008
10.1126/science.1161566

PirB is a Functional Receptor for Myelin Inhibitors of Axonal Regeneration

Jasvinder K. Atwal,[1*] Julie Pinkston-Gosse,[1*] Josh Syken,[2] Scott Stawicki,[3] Yan Wu,[3] Carla Shatz,[2,4] Marc Tessier-Lavigne[1]†

A major barrier to regenerating axons after injury in the mammalian central nervous system is an unfavorable milieu. Three proteins found in myelin—Nogo, MAG, and OMgp—inhibit axon regeneration in vitro and bind to the glycosylphosphatidylinositol-anchored Nogo receptor (NgR). However, genetic deletion of NgR has only a modest disinhibitory effect, suggesting that other binding receptors for these molecules probably exist. With the use of expression cloning, we have found that paired immunoglobulin-like receptor B (PirB), which has been implicated in nervous system plasticity, is a high-affinity receptor for Nogo, MAG, and OMgp. Interfering with PirB activity, either with antibodies or genetically, partially rescues neurite inhibition by Nogo66, MAG, OMgp, and myelin in cultured neurons. Blocking both PirB and NgR activities leads to near-complete release from myelin inhibition. Our results implicate PirB in mediating regeneration block, identify PirB as a potential target for axon regeneration therapies, and provide an explanation for the similar enhancements of visual system plasticity in PirB and NgR knockout mice.

Myelin, an insulating layer surrounding axons, is thought to pose an obstacle to axon regeneration, inhibiting neurite outgrowth in vitro and contributing to regeneration failure in vivo. The NgR, a candidate receptor for the myelin-derived inhibitors Nogo, MAG, and OMgp (*1*–*5*), appears to be required for the acute inhibitory activity of these proteins, because genetic removal of NgR blocks acute growth-cone collapse in response to these factors when added in solution (*6*, *7*). However, genetic deletion of NgR does not relieve the chronic inhibition of neurite outgrowth by myelin inhibitors presented as substrates (*7*, *8*). Furthermore, genetic deletion of NgR does not enhance regeneration of corticospinal tract (CST) axons after dorsal hemisection (*6*, *8*), although some regeneration of raphespinal and rubrospinal tracts after spinal cord injury has been reported (*6*). These data suggest that NgR is important for mediating some of the inhibitory activity of myelin inhibitors but that other binding receptors for these factors remain to be identified. Such putative receptors could work either independently or in concert with NgR.

[1]Neurodegeneration Labs and Research Drug Discovery, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA. [2]Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA. [3]Department of Antibody Engineering, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA. [4]Bio-X Program, Stanford University, James H. Clark Center, 318 Campus Drive, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: marctl@gene.com

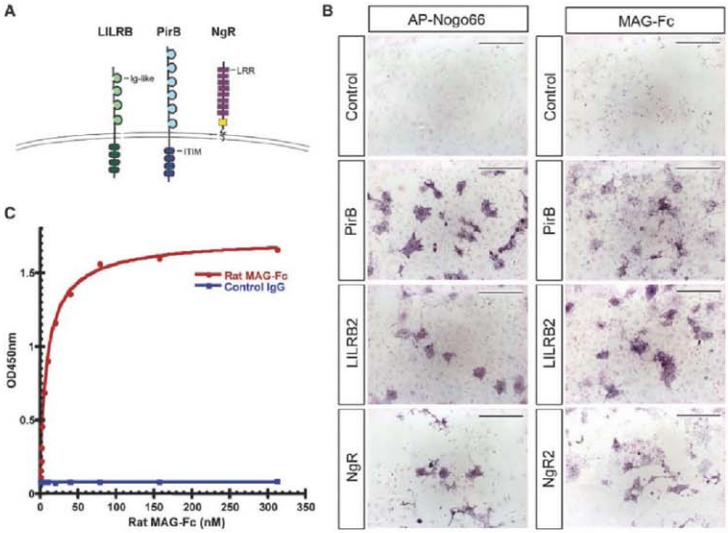


Fig. 1. LILRB2 and PirB can bind Nogo66 and MAG. **(A)** Schematic diagram of LILRB, PirB, and NgR receptors. ITIM, immunoreceptor tyrosine-based inhibitory motif; LRR, leucine-rich repeat. **(B)** COS cells were transfected with empty vector (control) or PirB or LILRB2 cDNA. After 48 hours, cells were incubated with AP-Nogo66 or MAG-Fc, and bound ligand was detected. NgR and NgR2 were used as positive controls for AP-Nogo66 and MAG-Fc binding, respectively. Scale bars, 200 μm. **(C)** The affinity of the MAG-Fc-PirB interaction is shown by one representative enzyme-linked immunosorbent assay binding curve. The experiment was repeated twice with similar results.

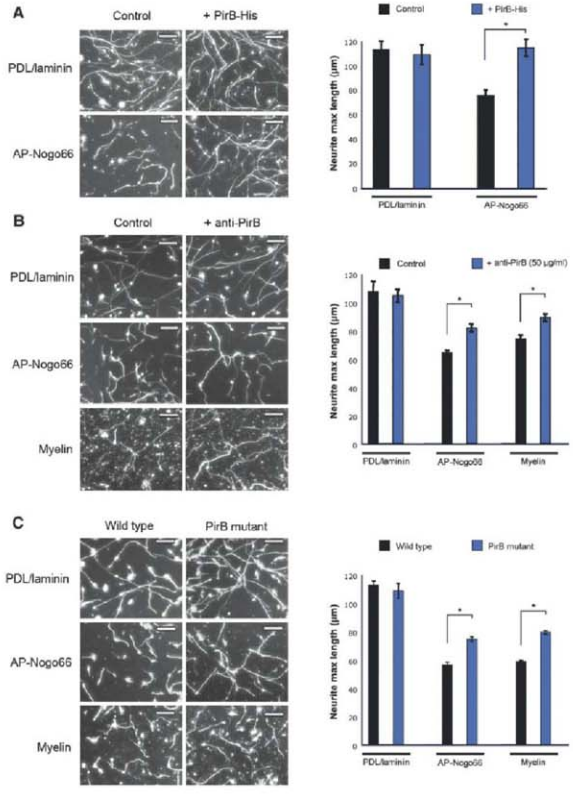


Fig. 2. Blocking PirB reverses inhibition of CGN outgrowth on AP-Nogo66 or myelin. Dissociated mouse P7 CGNs were plated on PDL/laminin (control), AP-Nogo66, or myelin to test inhibition by these substrates after various manipulations. In each panel, representative photomicrographs are shown on the left, and a graph measuring average neurite length (±SE, error bars) from one representative experiment is shown on the right. **(A)** Neurons were grown on PDL/laminin or AP-Nogo66, either alone or mixed with a fivefold excess of PirB extracellular domain (PirB-His). Neuronal inhibition by AP-Nogo66 was largely rescued by the presence of PirB-His. **(B)** Neurons grown on PDL/laminin, AP-Nogo66, or myelin were cultured in the presence or absence of anti-PirB.1 (50 μg/ml), which significantly reduced inhibition by either substrate. **(C)** Neurons cultured from either WT control mice or PirBTM mice were grown on PDL/laminin, AP-Nogo66, or myelin. PirBTM neurons were significantly less inhibited on either substrate (Student's *t* test,*$P < 0.01$; $n = 6$ wells per condition). Scale bars, 50 μm.

To identify candidate receptors for myelin inhibitors, we used expression cloning to screen an arrayed library of human cDNA pools (*9*). Our screen identified only two receptors for Nogo66: (i) NgR and (ii) the human leukocyte immunoglobulin (Ig)–like receptor B2 (LILRB2). LILRB2 is part of the B type subfamily of LILR receptors, which consists of five highly homologous family members in humans (*10*) (fig. S1). In mice, however, there is a single ortholog, paired immunoglobulin-like receptor B (PirB) (*11*). PirB shares only ~50% amino acid similarity with LILRB2, and it contains six Ig-like repeats, as opposed to four Ig-like repeats in LILRB proteins (Fig. 1A). Despite this low level of homology, we found that alkaline phosphatase (AP)–Nogo66 can bind PirB (Fig. 1B). The affinity of the AP-Nogo66-PirB interaction was similar to that of the AP-Nogo66 interaction with NgR (fig. S2). Given the unusual promiscuity of NgR binding to Nogo66, MAG, and OMgp, it is possible that LILRB2 and PirB also bind other myelin inhibitors. Indeed, we found that MAG-Fc and AP-OMgp bind PirB (Fig. 1B and fig. S2). The MAG-Fc-PirB interaction is high-affinity (half-maximal saturation of the interaction between purified MAG-Fc and purified PirB ectodomain: 13.8 ± 6 nM) (Fig. 1C).

To address whether PirB is a functional receptor for Nogo66, we focused on juvenile (P7) cerebellar granule neurons (CGNs), whose neurite outgrowth is inhibited when grown on AP-Nogo66 (*12*). Adult CGNs have been shown to express *PirB* (*13*), and we found that is also the case for juvenile CGNs, as assessed by reverse transcription polymerase chain reaction and in situ hybridization (fig. S3). We first tested the ability of a soluble ectodomain of PirB (PirB-His) to interfere with AP-Nogo66 inhibition in vitro. AP-Nogo66 inhibits neurite outgrowth of P7 CGNs to ~66% of untreated control levels (Fig. 2A). Inclusion of PirB-His in this assay reversed AP-Nogo66 inhibition, with neurite outgrowth returning essentially to control levels. These results are similar to those reported using the ectodomain of NgR to block inhibition by Nogo66 (*8*, *14*, *15*), and they indicate that PirB can bind the functionally inhibitory domain of Nogo66

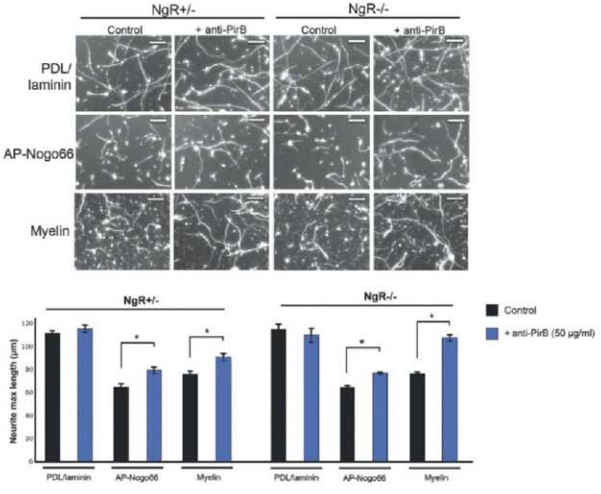


Fig. 3. PirB and NgR together contribute to myelin inhibition. Neurons were cultured on PDL/laminin, AP-Nogo66, or myelin substrates, and the PirB and NgR receptors were functionally blocked, either independently or in combination, to assess the contribution of these two pathways to inhibition. Representative photomicrographs are shown (top), and a graph measuring average neurite length (±SE, error bars) from one representative experiment is shown (bottom). NgR functional blockade was achieved by culturing neurons from NgR-null mice. PirB blockade was achieved by culturing neurons in the presence of anti-PirB.1 (50 μg/ml) (Student's t test,*$P < 0.01$; $n = 6$ per condition). Scale bars, 50 μm.

but do not address whether endogenous PirB in CGNs mediates inhibition by AP-Nogo66. Therefore, antibodies to PirB that are capable of interfering with the PirB-Nogo66 interaction were generated. Using a phage display platform (*16*) directed against the extracellular domain of PirB, we screened multiple clones for their ability to block binding of AP-Nogo66 to PirB (fig. S4). Clone YW259.2 (hereafter referred to as anti-PirB.1), which interfered best with AP-Nogo66-PirB binding, had a dissociation constant K_d of 5 nM for PirB. Anti-PirB.1 had no effect on the baseline axon growth of CGNs. However, it significantly reduced inhibition by AP-Nogo66 or myelin in cultured CGNs (Fig. 2B), rescuing neurite outgrowth to 59% from 41% on AP-Nogo66 and to 62% from 47% on myelin. Similar results were seen with MAG as an inhibitory substrate or with a different neuronal cell type [dorsal root ganglion (DRG) neurons] (fig. S5). These results suggest that PirB is a functional receptor mediating long-term inhibition of neurite outgrowth.

To confirm this result, we made use of mice carrying a loss-of-function *PirB* allele, the PirBTM mice, in which four exons encoding the transmembrane domain and part of the PirB intracellular domain have been removed (*13*). CGNs were cultured from PirBTM mice or wild-type (WT) littermates on a control substrate, AP-Nogo66, or myelin. On the control substrate (PDL/laminin), PirBTM neurons behaved similarly to WT neurons (Fig. 2C). However, neurite outgrowth from PirBTM neurons was markedly less inhibited than that from WT neurons on either AP-Nogo66 or myelin. On AP-Nogo66, outgrowth from WT neurons was inhibited to 50% of control levels, whereas PirBTM neurons were inhibited to 66%. Similarly, on myelin, WT neurons were inhibited to 52% of control levels, whereas PirBTM neurons were inhibited to 70%. Again, we observed similar partial disinhibition of PirBTM DRG neurons on both myelin and AP-Nogo66 (fig. S5). We also saw disinhibition of PirBTM CGNs on MAG and OMgp (fig. S5). These findings indicate that PirB is indeed a functional receptor for AP-Nogo66, MAG, OMgp, and myelin-mediated inhibition of neurite growth. However, loss of PirB activity does not fully rescue outgrowth.

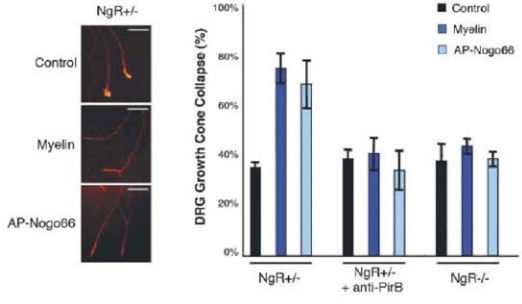


Fig. 4. Both PirB and NgR are required to mediate growth-cone collapse by myelin inhibitors. Growth cones of postnatal DRG axons were treated with medium alone (control), myelin (3 μg/ml), or AP-Nogo66 (100 nM) for 30 min to stimulate collapse and were stained with rhodamine-phalloidin to visualize growth cones. Representative photomicrographs are shown on the left, and a graph measuring percent of growth-cone collapse (±SE, error bars) from cumulative experiments is shown on the right ($n \geq 4$ per condition). Scale bars, 50 μm.

To address this concept, both PirB and NgR function were blocked together in CGNs by culturing neurons from NgR-null mice in the presence

of anti-PirB.1. As we have reported previously (*8*), $NgR^{-/-}$ CGN neurite outgrowth is inhibited by AP-Nogo66 or myelin to the same extent as that in WT neurons (50% and 49%) (Fig. 3). Anti-PirB.1 antibody treatment of $NgR^{-/-}$ neurons partially reversed inhibition by either AP-Nogo66 or myelin, as discussed above with WT neurons. Anti-PirB.1 treatment of $NgR^{-/-}$ neurons resulted in a similar partial disinhibition on AP-Nogo66 but did not provide any further rescue, suggesting that NgR is not involved in AP-Nogo66 inhibition. In contrast, anti-PirB.1 treatment of $NgR^{-/-}$ neurons restored neurite outgrowth on myelin to nearly control levels. Thus, it appears that PirB, but not NgR, is required for substrate inhibition by AP-Nogo66 in CGNs, but only accounts for it partly. In contrast, PirB and NgR both contribute to the substrate inhibition imparted by myelin.

Because NgR is required for growth-cone collapse in response to various myelin inhibitors (*6*, *7*), it is possible that PirB is also involved in this more acute response. For this experiment, we used sensory neurons from the DRG of 3-week-old mice, confirmed to express PirB (fig. S3). Similar to what has been shown by others (*6*, *7*), we found that growth cones in this culture system have a high baseline level of collapse (~30%), which is further increased (to ~75%) by incubation with AP-Nogo66 or myelin (Fig. 4). As reported previously (*6*, *7*), this collapse was largely abolished in $NgR^{-/-}$ neurons. In addition, blocking PirB function with anti-PirB.1 was also sufficient to reverse growth-cone collapse by these inhibitors.

Together, these results support a previously unknown role for PirB as a necessary receptor for neurite inhibition by myelin extracts and, more specifically, by the myelin-associated inhibitors Nogo66, MAG, and OMgp. PirB appears to be a more substantial mediator of substrate inhibition than NgR, as removal of PirB function alone (either genetically or with antibodies) partially disinhibits growth on both myelin extracts and inhibitors, whereas genetic removal of NgR alone does not disinhibit on any of these substrates. However, NgR appears to play an adjunct role in mediating inhibition by myelin extracts (but not Nogo66), because genetic removal of NgR can augment the disinhibition caused by anti-PirB antibodies on myelin (but not on Nogo66). Other co-receptors or modulators may also contribute in parallel, such as p75, TROY, LINGO, and gangliosides (*17*, *18*). Conversely, other PirB ligands, including major histocompatibility complex class I proteins, may contribute to the inhibitory action of myelin (*13*, *19*, *20*). Our finding of collaboration between PirB and NgR may help to explain the surprising lack of enhanced CST regeneration after dorsal spinal cord hemisection in NgR knockout mice (*6*, *8*), despite the reported regeneration or sprouting seen in rodents infused with the NgR ectodomain (*21*). Thus, it might be necessary to remove both PirB and NgR to achieve extensive regeneration in vivo. In addition, because on Nogo66 substrate the genetic removal of NgR does not further augment the partial disinhibitory effect of PirB removal, it is likely that there are additional binding receptors for Nogo66.

Although PirB appears to be a more important receptor for substrate inhibition than NgR, inactivation of either PirB or NgR alone is sufficient to block the acute growth-cone collapse caused by the addition of myelin inhibitors. This observation suggests that collapse is a more demanding process, requiring both PirB and NgR activities, acting either in parallel or together. In this context, it is of interest that PirB and NgR receptors have recently been shown to play similar roles in limiting plasticity of synaptic connections in the visual cortex. In mice lacking either receptor, eye closure during a critical developmental period results in excessive strengthening of connections via the open eye (*13*, *22*). The mechanisms responsible for the effect of both receptors in mediating growth-cone collapse could also underlie the commonality of their role in ocular dominance plasticity.

The mechanism by which PirB signals to inhibit axon growth in response to myelin inhibitors is not clear. However, PirB has been shown to antagonize the function of integrin receptors (*23*) and to recruit both src homology 2–containing protein tyrosine phosphatase (SHP)–1 and SHP-2 phosphatases (*13*, *24*); either or both of these events could attenuate normal neurite outgrowth. In humans, one or more members of the LILRB gene family might also play a role in regeneration. The blockade of PirB/LILRB activity, either with antibodies or by other means, provides an important target for therapeutic interventions to stimulate axonal regeneration.

References and Notes

1. A. E. Fournier, T. GrandPre, S. M. Strittmatter, *Nature* **409**, 341 (2001).
2. M. Domeniconi *et al.*, *Neuron* **35**, 283 (2002).
3. B. P. Liu, A. Fournier, T. GrandPre, S. M. Strittmatter, *Science* **297**, 1190 (2002); published online 27 June 2002 (10.1126/science.1073031).
4. K. C. Wang *et al.*, *Nature* **417**, 941 (2002).
5. J. K. Huang *et al.*, *Science* **310**, 1813 (2005); published online 16 November 2005 (10.1126/science.1118313).
6. J. E. Kim, B. P. Liu, J. H. Park, S. M. Strittmatter, *Neuron* **44**, 439 (2004).
7. O. Chivatakarn, S. Kaneko, Z. He, M. Tessier-Lavigne, R. J. Giger, *J. Neurosci.* **27**, 7117 (2007).
8. B. Zheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1205 (2005).
9. Materials and methods are available as supporting material on *Science* Online.
10. A. M. Martin, J. K. Kuiski, C. Witt, P. Pontarotti, F. T. Christiansen, *Trends Immunol.* **23**, 81 (2002).
11. H. Kubagawa, P. D. Burrows, M. D. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5261 (1997).
12. K. C. Wang, J. A. Kim, R. Sivasankaran, R. Segal, Z. He, *Nature* **420**, 74 (2002).
13. J. Syken, T. GrandPre, P. O. Kanold, C. J. Shatz, *Science* **313**, 1795 (2006); published online 16 August 2006 (10.1126/science.1128232).
14. A. E. Fournier, G. C. Gould, B. P. Liu, S. M. Strittmatter, *J. Neurosci.* **22**, 8876 (2002).
15. X. L. He *et al.*, *Neuron* **38**, 177 (2003).
16. W. C. Liang *et al.*, *J. Mol. Biol.* **366**, 815 (2007).
17. A. R. Walmsley, A. K. Mir, *Curr. Pharm. Des.* **13**, 2470 (2007).
18. N. R. Mehta, P. H. Lopez, A. A. Vyas, R. L. Schnaar, *J. Biol. Chem.* **282**, 27875 (2007).
19. M. Colonna *et al.*, *J. Exp. Med.* **186**, 1809 (1997).
20. A. Nakamura, E. Kobayashi, T. Takai, *Nat. Immunol.* **5**, 623 (2004).
21. S. Li *et al.*, *J. Neurosci.* **24**, 10511 (2004).
22. A. W. McGee, Y. Yang, Q. S. Fischer, N. W. Daw, S. M. Strittmatter, *Science* **309**, 2222 (2005).
23. S. Pereira, H. Zhang, T. Takai, C. A. Lowell, *J. Immunol.* **173**, 5757 (2004).
24. B. Chen, L. Hammonds-Odie, J. Perron, B. A. Masters, J. L. Bixby, *Dev. Biol.* **252**, 170 (2002).

25. We thank J. Brady and M. Yan for assistance with expression cloning and members of the Tessier-Lavigne and Shatz laboratories for helpful suggestions and discussions. This work was supported by Genentech.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/967/DC1
Materials and Methods
Figs. S1 to S5
References

30 May 2008; accepted 11 September 2008
10.1126/science.1161151

"Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech

Elia Formisano,* Federico De Martino, Milene Bonte, Rainer Goebel

Can we decipher speech content ("what" is being said) and speaker identity ("who" is saying it) from observations of brain activity of a listener? Here, we combine functional magnetic resonance imaging with a data-mining algorithm and retrieve what and whom a person is listening to from the neural fingerprints that speech and voice signals elicit in the listener's auditory cortex. These cortical fingerprints are spatially distributed and insensitive to acoustic variations of the input so as to permit the brain-based recognition of learned speech from unknown speakers and of learned voices from previously unheard utterances. Our findings unravel the detailed cortical layout and computational properties of the neural populations at the basis of human speech recognition and speaker identification.

In everyday life, we automatically and effortlessly decode speech into language independently of who speaks. Similarly, we recognize a speaker's voice independently of what she or he says. Cognitive and connectionist models postulate that this efficiency depends on the ability of our speech perception and speaker identification systems to extract relevant features from the sen-

sory input and to form efficient abstract representations (*1–3*). These representations are invariant to changes of the acoustic input, which ensures efficient processing and confers a high robustness to noise or to signal distortion. Relevant psycholinguist models consider abstract entities such as phonemes as the building blocks of the computational chain that transforms an acoustic waveform into a meaningful concept (*2*, *3*). There is also psychoacoustic evidence that the identification of a speaker relies on the extraction of invariant paralinguistic features of his/her voice, such as fundamental frequency (*1*).

Numerous functional neuroimaging studies have provided important insights on the cortical organization of speech (*4–11*) and voice (*12*, *13*) processing. However, the subtraction-based experimental logic and the limited neuroanatomical detail only allow for partial and indirect inferences on what distinguishes the auditory cortical representations of two natural speech or vocal sounds. Furthermore, it remains unclear how a speech sound is transformed into the more abstract entity of "phoneme" or "speaker" identity. Beyond subtraction, results from functional magnetic resonance adaptation suggest the involvement in voice identification of a specialized region in the right anterior superior temporal sulcus (STS) (*14*). For speech processing, a hierarchical fractioning of cortical regions for sound-based and a more abstract higher-level processing has been suggested [(*15*) and supporting online text].

In the present study, we investigate speech and voice recognition and abstraction at the level of representation and processing of individual sounds. By combining multivariate statistical pattern recognition with single-trial functional magnetic resonance imaging (fMRI) (*16–20*), we estimate and decode the distinct activation patterns elicited by different speech sounds and directly assess the invariance of the estimated neural representations to acoustic variations of the sensory input.

High spatial resolution (1.5 mm × 1.5 mm × 2 mm) functional images of the auditory cortex were collected while participants ($n = 7$) listened to speech sounds consisting of three Dutch vowels (/a/, /i/, /u/) recorded from three native Dutch speakers (Fig. 1) (*21*). Consistent with previous studies (*4–6*, *8*, *10*, *12*, *14*, *15*, *22*), all sounds evoked significant fMRI responses in a wide expanse of the superior temporal cortex, including early auditory areas (Heschl's gyrus) and multiple regions in the planum temporale (PT), along the superior temporal gyrus, the STS, and the middle temporal gyrus. Univariate statistical contrasts, however, yielded only weak response differences (below significance) or no differences between conditions (fig. S2)

After this initial analysis, we asked whether the estimation of a multivoxel activation fingerprint of a sound would allow deciphering its content and the identity of the speaker. With a method based on a machine learning classification algorithm (support vector machine) and recursive feature elimination (*23*, *24*), we performed two complementary analyses. We labeled the stimuli and corresponding response patterns either according to the vowel dimension irrespective of the speaker dimension ("vowel learning") or according to the speaker dimension irrespective of the vowel dimension ("speaker learning"). This led to the grouping of stimuli and responses in the three conditions: /a/, /i/, and /u/ and sp1, sp2, and sp3, respectively. We then examined whether our algorithm, after being trained with a subset of labeled responses (50 trials), would correctly classify the remaining unlabeled responses (10 trials). In all subjects and in all possible pairwise comparisons, the algorithm successfully learned the functional relation between sounds and corresponding spatial patterns and classified correctly the unlabeled sound-evoked patterns, both in the case of vowel learning [/a/ versus /i/ = 0.65 (mean correctness), $P = 6 \times 10^{-5}$; /a/ versus /u/ = 0.69, $P = 2 \times 10^{-5}$; /i/ versus /u/ = 0.63, $P = 4 \times 10^{-4}$ (Fig. 2A)]; and speaker learning, [sp1 versus sp2 = 0.70, $P = 3 \times 10^{-5}$; sp1 versus sp3 = 0.67, $P = 8 \times 10^{-5}$; sp2 versus sp3 = 0.62, $P = 2 \times 10^{-5}$ (Fig. 2B)].

To investigate layout and consistency across subjects of the spatial patterns that make this decoding possible, we generated group discriminative maps (Fig. 3 and fig. S3), i.e., maps of the cortical locations that contribute most to the discrimination of conditions. Single-subject reliability maps are reported in fig. S4. Discriminative patterns for vowels [red color (Fig. 3B and fig. S3)] were widely distributed bilaterally in the superior temporal cortex and included regions in

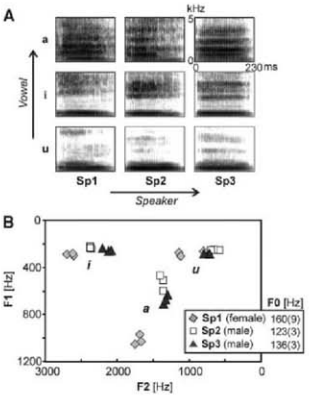


Fig. 1. Experimental design and stimuli. **(A)** Example of spectrograms of the stimuli from the nine conditions (three vowels × three speakers). Stimuli were presented during the silent intervals of fMRI measurements and were three natural Dutch vowels (/a/, /i/, and /u/) spoken by three native Dutch speakers (sp1: female, sp2: male, and sp3: male). **(B)** Representation of the vowels based on the first two formants (F1, F2). Each of the conditions was formed by grouping three different utterances from the same speaker. The insert indicates mean value and standard deviation of the fundamental frequency (F0) for each of the speakers.

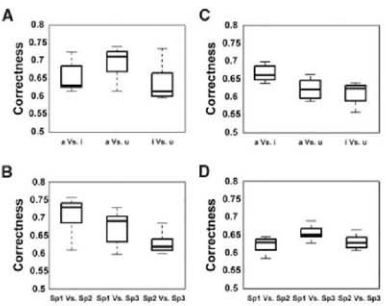


Fig. 2. Performance of the brain-based decoding of vowels and speakers. Correctness (median value and distribution) of all pairwise classifications when training and testing of the algorithm were based on different subsets of responses to the same stimuli [**(A)** vowels; **(B)** speakers] or when training and testing were based on responses to different speakers **(C)** and vowels **(D)** (chance level is 0.5).

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, University of Maastricht, 6200 MD Maastricht, Netherlands.

*To whom correspondence should be addressed. E-mail: e.formisano@psychology.unimaas.nl

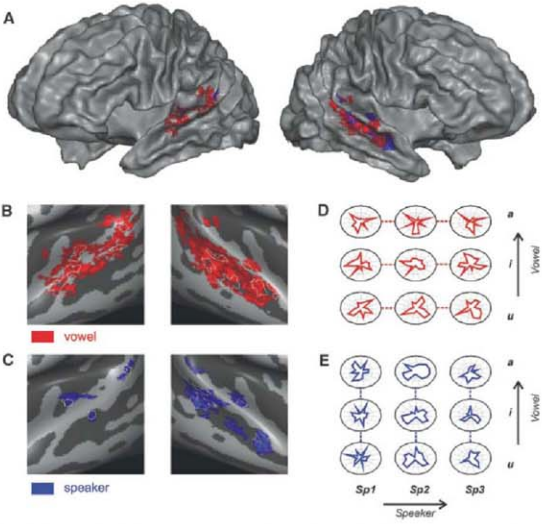


Fig. 3. Cortical discriminative maps and activation fingerprints for decoding of vowels and speakers. **(A to C)** Group discriminative maps obtained from cortex-based realignment of individual maps. Maps are visualized on the folded (A) or inflated representation of the cortex [auditory cortex detail in (B and C); light gray, gyri, dark gray, sulci] resulting from the realignment of the cortices of the seven participants. A location was color-coded (vowels, red; speakers, blue) if it was present on the individual maps of at least four of the seven subjects. This corresponds to a false discovery rate–corrected threshold of $q = 6 \times 10^{-4}$ for vowels and $q = 9 \times 10^{-4}$ for speakers (*21*). Outlined regions in (B) and (C) indicate cortical regions that were also included in the group maps obtained with the generalization analysis. **(D and E)** Activation fingerprints of the sounds created from the 15 most discriminative voxels for decoding of vowels (D) and speakers (E) (single-subject data, subject 1). Each axis of the polar plot forming a fingerprint displays the normalized activation level in a voxel. Note the similarity among the fingerprints of the same vowel [horizontal direction in (D)] or speaker [vertical direction in (E)].

the anterior-lateral portion of Heschl's gyrus–Heschl's sulcus, in the PT (mainly in the left hemisphere), and extended portions of the STS/STG (both hemispheres). Discriminative patterns for speakers [blue color (Fig. 3C and fig. S3)] were more confined and right-lateralized than those obtained for vowel discrimination. These patterns included the lateral portion of Heschl's gyrus–Heschl's sulcus, located in the posterior adjacenty to a similar region described for vowel discrimination and three clustered regions along the anterior-posterior axis of the right STS, also interspersed with vowels regions (fig. S3). These findings indicate a spatially distributed model for both the representation of vowel and speaker identity (see supporting online text).

Encouraged by these results, we tested the capability of our algorithm to decipher the brain activity into speech content and speaker identity also in the case of completely novel stimuli (i.e., stimuli not used during the training). We trained the algorithm in discriminating vowels with samples from one speaker (e.g., /a/ versus /i/ for sp1) or one vowel (e.g., sp1 versus sp2 for /a/) and tested the correctness of this discrimination in the other speakers (e.g., sp2 and sp3) or in the other vowels (e.g., /i/ and /u/). With this strategy, stimuli used for training and for testing differ in many acoustical dimensions. An accurate decoding of activation patterns associated with the test stimuli would thus indicate that the learned functional relation between a cortical activation pattern and a vowel (or a speaker) entails information on that vowel (or speaker) beyond the contingent mapping of its acoustic properties. Despite the small number of training samples (20 trials), classification of novel stimuli was accurate in all subjects and in all possible pairwise comparisons, both in the case of vowels [/a/ versus /i/ = 0.66 (mean accuracy), $P = 1 \times 10^{-6}$; /a/ versus /u/ = 0.62, $P = 3 \times 10^{-5}$; /i/ versus /u/ = 0.60, $P = 7 \times 10^{-5}$ (Fig. 2C)] and in the case of speakers [sp1 versus sp2 = 0.62 (mean accuracy), $P = 6 \times 10^{-6}$; sp1 versus sp3 = 0.65, $P = 8 \times 10^{-7}$; sp2 versus sp3 = 0.63, $P = 2 \times 10^{-6}$ (Fig. 2D)]. Although sparser, the corresponding discriminative maps included a subset of the locations highlighted by the previous analyses (outlined regions in Fig. 3, B and C).

The abstract nature of the estimated cortical representations is illustrated in Figs. 3, D and E, and 4. First, we visualized the speaker-invariant cortical representation of vowels and the vowel-invariant representation of speakers for which we used multidimensional displays [fingerprints (Fig. 3, D and E)]. Second, we visualized the relation between the discriminative patterns of activations for the nine conditions using self organizing maps (SOMs) (*21*), which convert complex relations between high-dimensional items into simple geometric relations. The spatial proximity and grouping of the conditions in the SOM-based two-dimensional display thus reflects the level of abstraction and categorical information entailed in the discriminative patterns of vowels (Fig. 4, A and B) and speakers (Fig. 4, C and D). To investigate which acoustic features in the original sounds drive this neural abstraction, we examined the relative distance between the brain-based representations of the stimuli and their description in terms of typical acoustic features (formants, Fig. 1B). We found that the distances between the cortical representations of the sounds correlated best with a description of the stimulus based on the first two formants (F1, F2) in the case of vowels [$r = 0.75$, $P = 2 \times 10^{-7}$ (Fig. 4E and fig. S5)] and on the fundamental frequency (F0) in the case of speakers [$r = 0.64$, $p = 2 \times 10^{-5}$ (Fig. 4F and fig. S5)]. These results provide empirical support for cognitive models of speech and voice processing postulating the existence of intermediate computational entities resulting from the transformation of relevant acoustic features [the (F1, F2) pair for vowels and (F0) for speakers] and the suppression of the irrelevant ones.

Our findings demonstrate that an abstract representation of a vowel or speaker emerges from the joint encoding of information occurring not only in specialized higher-level regions but also in auditory regions, which—because of their anatomical connectivity and response properties—have been associated with early stages of sound processing. This is in agreement with recent neurophysiological findings indicating that neurons in early auditory regions may exhibit complex spectrotemporal receptive fields and may participate in high-level encoding of auditory objects (*25–29*), e.g., via local feedback loops and reentrant processing. Taken together, these results prompt a revision of models of phoneme and voice abstraction, which assumes that a hierarchy of processing steps is "mapped" into a functional hierarchy of specialized neural modules.

In conclusion, we demonstrated the feasibility of decoding speech content and speaker

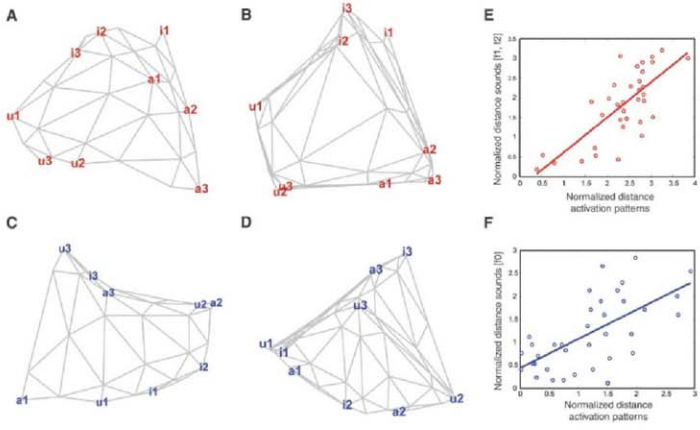


Fig. 4. Visualization of the brain-based representation of the sounds and relation with acoustical features. (**A** to **D**) SOM-based display of the discriminative patterns in a single-subject (subject 1, A and C) and in the group of seven subjects (B and D) for vowel (A and B) and speaker learning (C and D). (**E** and **F**) Relation between normalized distances of the multidimensional auditory cortical activation patterns and normalized distances of the vowels in the (F1, F2) space of formants (E) and of the speakers in the space of fundamental frequency (F0) (F).

identity from observation of auditory cortical activation patterns of the listeners. Our analyses provided a detailed empirical demonstration of how the human brain forms computationally efficient representations required for speech comprehension and speaker identification. Our experimental settings, however, were restricted to three vowels and three speakers; furthermore, all sounds were presented in isolation to obtain distinct fMRI activation patterns. Extension of these results to identify a word or concatenation of words in streams of longer speech segments, provides a compelling challenge and will contribute to create a general brain-based decoder of sounds in the context of real-life situations.

References and Notes

1. P. Belin, S. Fecteau, C. Bedard, *Trends Cogn. Sci.* **8**, 129 (2004).
2. J. L. McClelland, J. L. Elman, *Cognit. Psychol.* **18**, 1 (1986).
3. D. Norris, J. M. McQueen, *Psychol. Rev.* **115**, 357 (2008).
4. J. R. Binder *et al.*, *Cereb. Cortex* **10**, 512 (2000).
5. G. Hickok, D. Poeppel, *Nat. Rev. Neurosci.* **8**, 393 (2007).
6. E. Liebenthal, J. R. Binder, S. M. Spitzer, E. T. Possing, D. A. Medler, *Cereb. Cortex* **15**, 1621 (2005).
7. R. Näätänen *et al.*, *Nature* **385**, 432 (1997).
8. J. Obleser *et al.*, *Hum. Brain Mapp.* **27**, 562 (2006).
9. J. Obleser, A. Lahiri, C. Eulitz, *J. Cogn. Neurosci.* **16**, 31 (2004).
10. J. Obleser, J. Zimmermann, J. Van Meter, J. P. Rauschecker, *Cereb. Cortex* **17**, 2251 (2007).
11. S. K. Scott, I. S. Johnsrude, *Trends Neurosci.* **26**, 100 (2003).
12. P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, B. Pike, *Nature* **403**, 309 (2000).
13. J. D. Warren, S. K. Scott, C. J. Price, T. D. Griffiths, *Neuroimage* **31**, 1389 (2006).
14. P. Belin, R. J. Zatorre, *Neuroreport* **14**, 2105 (2003).
15. M. H. Davis, I. S. Johnsrude, *J. Neurosci.* **23**, 3423 (2003).
16. J. V. Haxby *et al.*, *Science* **293**, 2425 (2001).
17. J. D. Haynes, G. Rees, *Nat. Neurosci.* **8**, 686 (2005).
18. Y. Kamitani, F. Tong, *Nat. Neurosci.* **8**, 679 (2005).
19. K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, *Nature* **452**, 352 (2008).
20. N. Kriegeskorte, R. Goebel, P. Bandettini, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3863 (2006).
21. Material and Methods and supporting data are available on *Science* Online
22. R. Desai, E. Liebenthal, E. Waldron, J. R. Binder, *J. Cogn. Neurosci.* **20**, 1174 (2008).
23. F. De Martino *et al.*, *Neuroimage* **43**, 44 (2008).
24. E. Formisano, F. De Martino, G. Valente, *Magn. Reson. Imaging* **26**, 921 (2008).
25. C. T. Engineer *et al.*, *Nat. Neurosci.* **11**, 603 (2008).
26. N. Mesgarani, S. V. David, J. B. Fritz, S. A. Shamma, *J. Acoust. Soc. Am.* **123**, 899 (2008).
27. L. Nelken, *Curr. Opin. Neurobiol.* **14**, 474 (2004).
28. F. W. Ohl, H. Scheich, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9440 (1997).
29. X. Wang, T. Lu, R. K. Snider, L. Liang, *Nature* **435**, 341 (2005).

30. Work supported by Maastricht University and the Netherlands Organization for Scientific Research (NWO), Innovational Research Incentives Scheme Vidi grant 452-04-337 (E.F.) and Veni grant 451-07-002 (M.B.). We thank J. Obleser, U. Hasson, J.-D. Haynes, N. Kriegeskorte, J. Ashburner, and G. Valente for comments and discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/970/DC1
Materials and Methods
Supporting Text
Figs. S1 to S5
References

6 August 2008; accepted 9 October 2008
10.1126/science.1164318

MICROFLUIDICS:
BRINGING NEW THINGS TO LIFE SCIENCE

Most everything we know about fluids is based on macroscale observation. But things are different at the microscale, where "channels" can be built with two walls instead of four, turbulence doesn't exist, and fluids can flow side by side without mixing. For years, microfluidics has been the subject of hyperbole, with some speculating that laboratory process would one day be miniaturized onto a chip. Current commercial offerings leverage microfluidics' most obvious advantages: small sample volumes, rapid results, and lower costs. But as researchers have discovered, microfluidics also has a second benefit: enabling experiments that cannot otherwise be accomplished. **By Jeffrey M. Perkel**

As a new assistant professor, Adela Ben-Yakar found a way to use low-energy laser pulses to sever a single neuron in the millimeter-sized nematode, *Caenorhabditis elegans*. The technique is called femtosecond laser axotomy, and according to Ben-Yakar, it has been well received.

"We developed nanoaxotomy in 2004," says Ben-Yakar, of the **University of Texas in Austin**, "and at a *C. elegans* meeting in June there was a whole session on using it."

Unfortunately, the technique is far too unwieldy for large-scale genetics. Between picking up the animal, positioning it on the microscope, anesthetizing it, and the surgery itself, each worm takes about 10 minutes, says Ben-Yakar, who has a Ph.D. in mechanical engineering and a postdoc in applied physics.

She could have thrown more students at the project, like so much laboratory cannon fodder, but that would have occupied them for years, she suspected. "I thought maybe I could make a microfluidic device and do the work with far fewer people in much less time."

What she created was a so-called "lab-on-a-chip" device capable of capturing a worm, holding it steady during surgery, and shunting the animal to any of three recovery chambers. The process takes under a minute per animal, she says.

It is an enabling technology for Ben-Yakar, who says she now can screen for new genes and chemicals that might influence neuroregeneration in vivo. It also is a development that, in a sense, encapsulates the promise of microfluidics, a field that is about more than miniaturization, lower reagent volumes, and cost.

"The desire to use less sample is important and it is what got people excited," says Nate Cosper, marketing director at **Caliper Life Sciences**. "But really what's different is you can now do things you previously couldn't do."

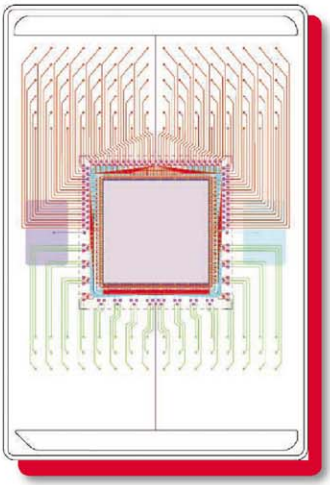
Thinking Small

"The trick with microfluidics is, you have to think small," says Glenn Walker, assistant professor of biomedical engineering at the **University of North Carolina** (UNC) and **North Carolina State University**. "You can't just take something that works big, shrink it down, and expect it to work."

That's because fluid dynamics don't scale, he says. Fluids that flow like water at the macroscopic level behave more like honey at the microscale. Inertial forces decline while shear forces become important. Turbulence becomes a nonissue, while interactions between the fluid and the channel wall become critical. Even the seemingly simplest of processes can become problematic. For instance, since magnetic forces also don't scale well, magnetic-based separations generally aren't as effective on a chip.

On the other hand, the physics of microfluidics is also enabling, Walker says. Two parallel fluid streams in a microchannel, if they flow fast enough, will continued >

CREDIT: IMAGE PROVIDED BY FLUIDIGM

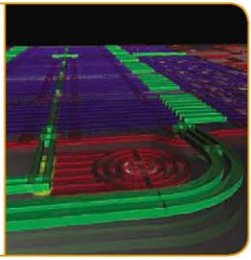


"The desire to use less sample is important and it is what got people excited. But really what's different is you can now do things you previously couldn't do."

Look for these **Upcoming** Articles
Cell Signaling — December 5
Proteomics: MS Purification/Separation — February 20

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, *nor is it meant to imply that their products or services are superior to those of other companies.*

"PDMS is just very easy to work with. There's a very simple procedure to go from design to master to mold, and you can seal the layers together almost effortlessly"



not mix, for example. "There will be a very sharp interface [between them]," Walker says. Thus, it is possible to bathe just one side of a cell with a chemical, to see how it reacts.

"That in particular is a really good example of a new, enabling tool that lets you do something you cannot do with pipettes and flasks," Walker says. "The ability to treat only part of a cell with a chemical is really powerful."

Yet there is something of a disconnect between microfluidics experts like Walker and the intended end-users of their engineering creativity. Despite a few notable commercial successes, such as Agilent Technologies' Bioanalyzer 2100, the majority of microfluidic advances remain bound up in the literature, beyond the reach of the life science community.

"The technology is in its early adolescence," says George Whitesides, professor of chemistry and chemical biology at **Harvard University**. "To me one of the big issues is—whether users are in analytical work or in biopharmaceuticals—that, quite correctly, they don't want to be in the business of making the device. Their expertise is not in microfabrication," Whitesides says. "So as with electronics, what happens is there is a critical mass of applications where it becomes worth it for manufacturers to get into the area, to make products that are commercially available."

"It's like a computer or a car: Most people cannot build one or fix one, but they can learn how to use them," says Stephen Quake, professor of bioengineering at **Stanford University** and the **Howard Hughes Medical Institute** and co-founder of microfluidics firm **Fluidigm**. "When a technology reaches that point, you can say it has arrived. And that's where we are right now for some applications."

Material Support

At least some of that success stems from advances in materials. While some build chips out of glass or hard plastic, many others (including Fluidigm) prefer poly(dimethylsiloxane) (PDMS), a rubbery material also used to make soft contact lenses. According to Whitesides, who heads one of two groups that first worked with the polymer, PDMS has become a popular choice for microfluidic device fabrication for its ease of use.

Glass and hard plastics are easy to break and hard to etch and bond, he says. In contrast, "PDMS is just very easy to work with. There's a very simple procedure to go from design to master to mold, and you can seal the layers together almost effortlessly," he says. A rubbery elastomer, PDMS won't break if dropped. It also breathes, so gases can exchange with the environment beyond the chip, while the material's springiness enables pneumatic control.

Pneumatic control circuits, which Quake developed while at the California Institute of Technology and licensed to Fluidigm, comprise two layers. Fluid flows through channels in the bottom layer. In the top layer, separated from the bottom by a thin layer of PDMS, empty control channels run perpendicular to the fluid channels. Pressurization of the control channels with air closes the fluid channels below, just as stepping on a hose will staunch the flow of water.

Ben-Yakar's nanoaxotomy chip uses this process to trap individual worms in a microfluidic "surgical suite" as if with a finger. For Fluidigm, pneumatic valves also enable system integration on an unprecedented level. The company's 96.96 Dynamic Array, for instance, uses more than 25,000 valves and about a million features to run 9,216 parallel 6.7-nl polymerase chain reactions in just three hours.

That's just miniaturization. But the company's newest chip is truly enabling, says Gajus Worthington, Fluidigm's co-founder, CEO, and president. "The [BioMark 12.765] Digital Array allows you to do things you could never do before. It literally does digital PCR."

"Suppose you have a 5-µl sample with a single molecule you want to test," he explains. "We break it into 1,000 different compartments, and we effectively increase the concentration of what you are looking for 1,000 times. That makes it 1,000 times easier to find."

That means it might be possible to, for instance, screen for evidence of fetal aneuploidy in maternal blood, a test that currently cannot otherwise be run, says Worthington.

Making the Impractical Practical

Swedish microfluidics firm **Cellectricon** targets a decidedly smaller niche with its Dynaflow system.

A microfluidic perfusion system intended for drug developers who use patch clamping for optimization of potential ion channel compounds, Dynaflow "is an add-on that enables the customer to rapidly control fluids around the patch-clamped cells," says Matthias Karlsson, chief technical officer and scientific co-founder. The cell is patch-clamped as normal and placed in the chip; then the system simply switches the fluid channel to which the cell happens to be exposed.

According to Karlsson, existing technology does not allow for controlled, rapid changes in the fluid environment around a clamped cell, and that can be a problem, because signaling events occur in milliseconds.

"If you cannot monitor with millisecond resolution, you will not see the reaction of the drug with the channel," he says, adding, "Dynaflow makes these experiments possible in an industrial setting."

According to Karlsson, "The benefit of having this on a microfluidic system is that the only mixing is by diffusion, not by turbulence, so the transition from one compound to another is very sharp, like a step function."

The throughput is about 200 compound wells per day, Karlsson says, or "10 to 20 times faster than what a skilled electrophysiologist could do now." The company's next-generation system, due late 2009, will bump that up to around 5,000 wells per day.

J. Michael Ramsey, Goldby Distinguished Professor of Chemistry at UNC and co-founder of Caliper Life Sciences, uses glass microfluidic devices to enable yet another highly impractical process: single-cell kinase assays.

Ramsey's collaborator, Nancy Allbritton, had previously developed a method to load cells with a fluorescent kinase substrate, lyse those

CREDIT: IMAGE PROVIDED BY FLUIDIGM.

cells one at a time using a laser, and then run capillary electrophoresis on the contents to measure enzymatic activity.

"They could do eight to 10 cells per day," says Ramsey. Using a microfluidic device built of glass, Ramsey's team has made the assay realistically useful. "We can do 10 cells per minute," he says, "and we could probably speed that up."

He explains, "We have a continuous flow of cells that stochastically arrive at a point where they are lysed, and the lysate is directed into a capillary electrophoresis channel and separated." It is the same assay Allbritton devised, except "it has essentially been automated."

Unique Advances

Sometimes microfluidic devices enable richer data from traditional assays. Caliper's Cosper says kinase assays run on the company's LabChip systems yield more than a number, for instance; because the reaction components flow through a separation channel, the data are both quantitative and qualitative.

"In a typical single-pot assay, you can get false positives or negatives because things interfere with the reaction and cause you to misinterpret the assay," he says. By running that same assay on the LabChip platform, "you get the same separation quality as in HPLC, to get very high quality data without artifacts. You cannot do that in a standard microwell."

Gyros has successfully commercialized a microlaboratory in compact disc format, utilizing flow-through immunoassay principles for protein quantification for the development of protein therapeutics. The Gyrolab system allows assays to be performed at nanoliter scale enabling users to extract more information from minute sample volumes, allowing repetitive sampling from individual animals, during the early development stages of protein therapeutics.

Gyrolab provides a completely different way of interacting analytes with capture reagent. "In our case, this is done under constant flow," says Mats Inganäs, Gyros' director of applications and technology assessment. "With ELISA, a sample needs to incubate in a static situation for hours to get the same reaction completed. Our system reduces both assay time—over a hundred assays can be processed in under one hour—and matrix effects."

New Takes on Old Techniques

David Beebe, professor of biomedical engineering at the **University of Wisconsin, Madison,** uses lab-on-a-chip technology to gain new insights into cell behavior.

Beebe's lab focuses on cell-cell communication, especially in cancer. One project studies the interaction between epithelial cells and the underlying stroma. "It's very difficult to do this with standard methods," he says, as these require either conditioned media or transwell systems, both of which will miss short-lived factors as well as reciprocal signaling events.

Instead, Beebe uses simple microfluidic alternatives to uncover novel phenomena not previously seen in a Petri dish. "For instance, we see a bimodal distribution of proliferative capacity [in primary mouse mammary cells]," he says, which he suspects reflect stem cells. "It is a cell behavior no one has observed in vitro before."

For Juan Santiago, associate professor of mechanical engineering at Stanford University, microfluidics also enables new approaches to PCR. He and his student, Alexandre Persat, are pursuing a process they call chemical-cycling PCR, which denatures DNA with a denaturant like formamide instead of heat. "Our chip works by cycling chemistry rather than temperature," Santiago explains. "We inject clouds of denaturant while holding the DNA stationary in an electric field."

Besides obviating heating and cooling elements (and thus, making PCR machines smaller and lighter), this approach is also faster than traditional PCR—Persat has amplified 194-bp targets in just 18 30-second cycles—lower-volume, and possibly more quantitative, Santiago says. In addition, because the amplified DNA is already in a microchannel, it can immediately be reacted or separated.

"As the PCR cycles occur, we are already separating the products," he says. "No one has ever done that before in real time." Santiago says the approach could find use in areas like forensics, where sample volume, sensitivity, and contamination are primary concerns.

Santiago says he "strictly" uses off-the-shelf chips, mostly available from Caliper for about $50, for his work. For those who want a more customized solution, PDMS fabrication is relatively straightforward, says Quake, and can easily be done in a lab, or even a garage, though a clean room is preferable. Stanford's microfluidics foundry (thebigone.stanford.edu/foundry/) can create microfluidics chips on demand.

Alternatively, you can try Mark Burns's approach. A professor of chemical engineering and biomedical engineering at the **University of Michigan,** Burns and student Minsoung Rhee developed a toolkit of 15 "microfluidic assembly blocks," essentially PDMS "LEGO" pieces that users can assemble into any desired structure.

"It's like making a mosaic," Burns explains. "You put each tile down, and it makes a picture when you're done. Only, instead of a picture, it is actually a functioning device."

An enabling technology for an enabling technology? Only time will tell.

Jeffrey Perkel is a freelance writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p0800029

Featured Participants

Caliper Life Sciences
www.caliperls.com

Cellectricon
www.cellectricon.com

Fluidigm
www.fluidigm.com

Gyros
www.gyros .com

Harvard University
www.harvard.edu

Howard Hughes Medical Institute
www.hhmi.org

North Carolina State University
www.ncsu.edu

Stanford University
www.stanford.edu

University of Michigan
www.umich.edu

University of North Carolina (UNC)
www.unc.edu

University of Texas, Austin
www.utexas.edu

University of Wisconsin, Madison
www.wisc.edu

Microfluidic Instruments

The LabChip GX and LabChip GXII benchtop microfluidics systems are for fast, automated, one-dimensional electrophoretic separations of protein, DNA, and RNA samples. The LabChip GX is a low-price, entry-level system targeted at genomics applications, while the GXII combines genomics and protein research applications. The instruments offer extended walk-away time, high throughput, and economical plate-processing ability. Both systems combine highly reproducible assay technology with advanced data analysis and visualization software. They offer 96-well and 384-well plate compatibility and the ability to select single wells at any location in a plate. Users can thoroughly analyze samples in seconds instead of minutes, eliminating throughput bottlenecks and improving efficiency.

Caliper Life Sciences
For information 877-522-2447
www.caliperLS.com

Microfluidic Chip

A reusable, high-performance liquid chromatography (HPLC), nanoflow-rate chip features titanium-dioxide-particle–based phosphopeptide enrichment. The chip is a multilayer polyimide laminate that contains an enrichment section with titanium dioxide beads flanked on both sides with C18 reversed-phase material. The three sections of the sandwich are separated by microfabricated frits. The unique sandwich configuration of the enrichment section provides researchers three modes of peptide analysis: standard peptide analysis, phosphopeptide analysis only, and combined peptide and phosphopeptide analysis. The chip can be used with Agilent HPLC-chip/mass spectrometry instrumentation, with a choice of mass spectrometers.

Agilent Technologies
For information 408-553-7191
www.agilent.com

Integrated Fluidic Circuit

The BioMark 96.96 Dynamic Array is a new integrated fluidic circuit (IFC) capable of performing 9,216 simultaneous real-time polymerase chain reaction (PCR) experiments running TaqMan assays in nanoliter quantities. This new-generation IFC enables life science researchers to achieve new levels of cost and logistical efficiency and flexibility, as well as comprehensive profiling from minuscule amounts of sample. The 96.96 Dynamic Array provides the flexibility of a microwell plate and the density of a microarray in one easy-to-use consumable. IFCs meter, multiplex, and combine nanoliter volumes of fluids thousands of times in parallel on a single chip with precise control and reproducibility. With this new dynamic array, researchers can set up 9,216 experiments with just 192 pipetting steps.

Fluidigm
For information 650-266-6081
www.fluidigm.com

Purification System

The compact Isolera Flash purification system enables chemists to achieve better separations in significantly shorter purification run times with new features that result in maximum purity and yield. The advanced thin layer chromatography (TLC)-to-gradient feature automatically creates elution gradients and suggests cartridge and sample size. Compounds can be collected on two separate wavelengths using a variable-wavelength detector option, resulting in increased compound recovery and cleaner fractions. Up to four solvents can be used in a single gradient to easily purify samples with diverse polarity. A third cosolvent can be isocratically pumped into the binary gradient, maintaining compound solubility and eliminating the over-pressure issues that can be caused by precipitating compounds. Users have full control over sample purification conditions, as a full range of editing options is available, including gradient, flow rate, collection volume, and fraction wavelengths and modes.

Biotage
For information 434-979-2319
www.biotage.com

Immunoassay Device

The Gravi ELISA-Cell is a USB-powered, microfluidics-based immunoassay device for running bead-based protocols with standard immunology reagents. The instrument works with Gravi ELISA-Chip (microchannel-harboring biosensors) for performing everything from sample preparation through signal detection. The compact device is designed for manual operation, to generate results wherever needed and most convenient. The device offers the flexibility of running almost any available enzyme-linked immunosorbent assay (ELISA) or affinity assay.

Calbiotech
For information 619-660-6162
www.calbiotech.com

Vibration Cancellation System

The TableTop PZT is an active, hard-mount vibration cancellation system for small precision instruments, particularly those in buildings where floor vibration is severe. The TableTop PZT features a lightweight, compact design; extended stroke piezoelectric actuators; and sub-HZ vibration cancellation, both vertical and horizontal. It has no soft air suspension. It has a payload capacity of up to 300 pounds.

Technical Manufacturing Corp.
For information 800-542-9725
www.techmfg.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.